UNIVERSITY OF ILLINOIS BULLETIN

ISSUED WEEKLY

Vol. XIX July 3, 1922 No. 45

[Entered as second-class matter December 11, 1912, at the post office at Urbana, Illinois, under the Act of August 24, 1912. Acceptance for mailing at the special rate of postage provided for in section 1103, Act of October 3, 1917, authorized July 31, 1918.]

BULLETIN NO. 10-17

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

by

BUREAU OF EDUCATIONAL RESEARCH

Prepared, in part, from a report by P. R. Stevenson,
former Assistant, Bureau of Educational Research.

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

BULLETIN NO. 10

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

by

BUREAU OF EDUCATIONAL RESEARCH

Prepared, in part, from a report by P. R. Stevenson,
former Assistant, Bureau of Educational Research.

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

TABLE OF CONTENTS

PREFACE

This investigation was initiated by Mr. B. R. Buckingham, formerly Director of the Bureau of Educational Research. It was planned and executed by him with the assistance of Mr. P. R. Stevenson, a full time assistant in the employ of the Bureau of Educational Research during the school year of 1920-21. The present Director of the Bureau of Educational Research had no connection with the study until late in the summer of 1920-21. The portion of this report which deals with "existing conditions in regard to class size" and "the relation of size of class in high schools to school efficiency" is based upon tabulations made by employees of the Bureau of Educational Research under Mr. Stevenson's direction and included by him in a report submitted to the present Director of the Bureau of Educational Research. The chapter devoted to the relation of the size of class in elementary schools to school efficiency is based upon tabulations made from the original data under the immediate direction of the present Director. The concept of the efficiency ratio and the use of this concept in the interpretation of the data are entirely the work of the present Director of the Bureau of Educational Research. The conclusions are also his own.

Certain limitations of the investigation, which the report discusses in detail, cause the results to have a limited practical significance, but it is thought that the publication is jusified for two reasons. In the first place, the concept of school efficiency and the analysis of the conditions which must be considered in any investigation relating to school efficiency should be helpful to future investigators not only of the question of class size but also of questions of other phases of school procedure. In the second place, the report emphasizes the need for careful planning which will result in the control of all factors involved in the teaching situation. There is also emphasis upon the need for securing normal conditions if the results are to be interpreted with reference to the modification of practise. Such analysis and careful thinking are not only important phases of educational research but they are the foundation upon which both the data collected and the statistical manipulation of them are based.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson, of the Chicago Public Schools, and of the school officials in certain other Illinois cities. Not only did they cooperate by permitting the collection of the data, but they actually made substantial contributions to the project by furnishing the test materials. The teachers in the schools concerned made a substantial contribution by scoring the tests and reporting them in a convenient form to the Bureau of Educational Research. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all those who have contributed to the project.

WALTER S. MONROE, *Director.*

May 26, 1922.

RELATION OF SIZE OF CLASS TO SCHOOL EFFICIENCY

CHAPTER I

INTRODUCTION: ANALYSIS OF PROBLEM

The problem. The problem of this investigation is to study the relation of the size of class to school efficiency, or what is the effect upon the efficiency of the school when the size of class is increased or decreased within certain limits.

Definition of terms: Class. In the high school, a class is defined as the number of pupils who are assigned to a single teacher for instruction during a single class period. In the elementary school, unless the instruction has been departmentalized, a class is the number of pupils assigned to a room over which a teacher has charge. For instructional purposes, a teacher in the elementary school may divide a class into two or three groups, but the total number of pupils receiving instruction from her is considered a class, as the term is used in this study.

School efficiency. Educators have borrowed the term "efficiency" from industry and business. In these fields, efficiency is expressed by a fraction whose maximum is 1.00. The numerator of this fraction, or "efficiency ratio," is the output, and the denominator is the input, or educational investment. In education, the output of a school system consists of the changes produced in the pupils, i. e., the controls of conduct that the school engenders. The educational output for a semester or a year is the total of all the changes that have been produced in the pupils during the period due to the influence of the school. The educational investment includes many factors, such as buildings, equipment, textbooks, teachers, supervision, and general administration. Although it is not imperative to do so, it is probably best to think of both the output and the investment as being expressed in terms of the average for one pupil.

We are accustomed to refer to the educational output as the achievements of the pupils. By means of educational tests and other instruments, we measure these achievements in terms of arbitrary units. Units of one type are implied in school marks. Other

units are defined by educational tests. In order to calculate the numerical value of the efficiency ratio, it would be necessary to determine the social value of the output in terms of dollars and cents or in terms of some unit which might be made common to both numerator and denominator. Obviously, we are not prepared to do this. However, we may use the ratio as a definition of school efficiency and inquire into the probable nature of the changes produced in it by the variations of certain factors upon which the educational output and educational investment depend. In making inferences concerning the fluctuations in the value of the efficiency ratio, it is necessary to remember that the production of certain achievements may be of little value when viewed in relation to our educational objectives. The attainment of certain levels of achievement may represent an educational output of considerable value, but advancement to higher levels may produce only slight increases in the value of the total output in this field. For example, the attainment of certain levels of ability in spelling has a distinct and relatively large social value, but advancement beyond these levels is accompanied by rapidly diminishing increments of value. Therefore, one must avoid the assumption that fluctuations in achievements are to be interpreted as having proportional values in terms of social worth.

Factors which affect school efficiency. For a given educational investment per pupil, the value of the efficiency ratio is affected when changes are made in the educational output. Methods of instruction, the plan of school organization, or the procedure of supervision may be modified even when there is no change in the investment. When modifications in the methods of using the investment result in changes in the educational output, there are resulting changes in the value of the efficiency ratio. On the other hand, it is possible that material changes may be made in the educational investment which are not accompanied by corresponding changes in the educational output. When this happens, the value of the efficiency ratio is changed, even though the actual educational output has remained constant.

In many cases, modifications in the method of using the educational investment are accompanied by changes in the magnitude of the educational investment as well as by changes in the educational output. Hence, we may have fluctuations occurring in both the numerator

and the denominator of the efficiency ratio. It is possible that these fluctuations may be connected in such a way that the value of the efficiency ratio remains constant, or it may be that its magnitude will vary. Because we are not able to calculate a numerical value of the efficiency ratio, a careful analysis is required to determine the probable changes in it when variations occur in the numerator and the denominator simultaneously.

The achievements of pupils are materially affected by their general intelligence or capacity to learn. Since individual pupils and also groups of pupils have been shown to exhibit marked individual differences when measured with respect to this trait, it is necessary to make due allowance for differences in general intelligence when comparing different school units with respect to efficiency. In case this is not done an error of interpretation will be made by attributing a higher degree of efficiency to those units which consist of pupils of superior general intelligence.

The effect of varying the size of class. The size of class is one item of the plan of the organization. From this point of view, it may be considered one of the methods of using the educational investment. Consequently, we may expect to find that changes in the size of class produce variations in the achievements of pupils. The size of the class is, also, one of the factors which determines the educational investment. In the elementary school, where a class means the number of pupils assigned to a teacher, the cost of instruction per pupil varies inversely with the size of class.[1] In a high school the size of class does not completely determine the number of student hours of instruction which a teacher gives, but it is a potent factor in this determination. In general, an increase in the size of class in the high school will tend to result in a marked decrease in the educational investment per pupil. Hence, in studying the effect of varying the size of class upon the efficiency of the school, it is necessary for us to inquire into the resulting changes in both the educational output and the educational investment. It is only when we have done this that we are in a position to make inferences concerning the effect of variations in the size of class upon the efficiency of the school.

[1]In making this statement, no account is taken of investments made in supervision, in instruction by special teachers, and in equipment.

The practical importance of a study of the relation of class size to school efficiency. During recent years, school administrators have faced the problem of providing instruction for a rapidly increasing enrollment and, at the same time, of meeting the demands from teachers for increased salaries. In meeting these two demands, there has been a tendency to increase the number of pupils instructed by a teacher in order to keep the total educational expenditure within the income of the school system. In the elementary school, and to a considerable extent in the high school, the number of pupils instructed by a teacher has been increased by increasing the size of classes. It is obvious that pupils in large classes have less opportunity for recitation and, in general, receive less individual attention from the teacher, both within and outside of the class period. Thus, the question has naturally been raised concerning the effect upon the efficiency of a school when the size of the class is increased. In the secondary school added emphasis is given to the question because certain accrediting agencies require that the size of the class not exceed a certain fixed maximum.

Connection between size of class and methods of instruction. It is necessary to bear in mind that, when the number of pupils instructed by a teacher is increased, there is a corresponding increase in the amount of work required of the teacher, unless there are compensating changes in the methods of instruction. For example, according to our present methods, it is customary to require a great deal of written work of pupils studying English composition. A teacher is expected to read with considerable care the compositions submitted by pupils and to provide a systematic procedure for correcting the errors made. Thus, an increase in the number of pupils to be taught increases the work required of the teacher unless the number of compositions required from each pupil is reduced or a different system of handling them is used. Much the same conditions prevail in a number of other subjects in which notebooks or other written work of some sort are customarily required. Similar statements can be made with reference to individual work with pupils outside of the regular class period. It is obvious that there is a limit to the amount of work which may legitimately be required of a teacher. When the optimal teaching load has been reached, any increase in the number of pupils assigned to a teacher should

be expected to be accompanied by compensating changes in the procedure of instruction. When such changes are made in the teaching procedure, they, as well as the size of the class, must be considered in respect to the effect upon the efficiency of the school.

Limitations of the present investigation. In this investigation of the effect of varying the size of class upon school efficiency, it was intended that all other factors which affect either the achievements of pupils or the educational investment should be kept constant, i. e., they should be the same for classes of different sizes. No information is available to show the extent to which this intention was realized except in the case of the quality of the pupil material. Group intelligence tests were used to secure equivalence of capacity to learn in the two groups of pupils. The possibility of a lack of equivalence of such factors as home environment, nationality, attitude toward school work, previous school experience with respect to size of class, time of day (applies only to high school), etc., makes it necessary to exercise due caution in interpreting the results of the investigation.

So far as the present writer is aware, the instruction that prevailed in the groups concerned in this investigation involved no unusual features, and the same methods of instruction were followed in the two types of classes. It is possible that the highest degree of efficiency for classes of a given size would be attained if the methods of instruction were selected with particular reference to this size of class. It would not be surprising to find that methods of instruction which were most effective in small classes would be considerably less effective with large classes, and that methods well suited to the handling of large classes would not give the best results when used with small classes. This possibility was not considered in this investigation. Hence, the findings should not be accepted as final. There is still need for an investigation in which methods of instruction are adapted to size of class.

Finally, it must be remembered that the problem of the size of class is not entirely a problem of the efficiency of the school. It is also a problem of the teacher. It is not humane and it is not socially profitable to assign teaching loads so heavy that teachers become overworked. The problem of the teacher was not considered in this investigation.

CHAPTER II

EXISTING CONDITIONS IN REGARD TO CLASS SIZE

In order that the significance of the two investigations to be reported in the following chapters may be more fully appreciated, certain facts are presented, concerning the size of class under present school conditions in Illinois.

Size of class in elementary schools in Illinois outside of Chicago. Data with reference to the size of class in the elementary schools of the state, outside of Chicago, were secured by sending a questionnaire to the superintendent of public schools in all cities and towns listed in the Illinois School Directory for 1920-21 as having six or more elementary teachers. The questionnaire asked for the number of elementary teachers having classes of the following sizes: less than 20, 20 to 29, 30 to 39, 40 to 49, and 50 and over. This information was requested for each of the school years of 1918-19, 1919-20, and 1920-21. Complete reports were received from 180 cities and towns. These are summarized in Table I. The total number of classes for which a report was secured varied from 9,422, 1918-19, to 10,403, in 1920-21. The median size of class varied from 41.4 pupils, 1918-19, to 43.3 pupils, in 1919-20. In 1920-21, the size of class was slightly less than that for the preceding year. This table also shows that, in the cities reporting, slightly more than 3 percent of all classes contained 50 or more pupils. On the other hand, between 7 and 8 percent of the classes had less than 30 pupils. The greatest change in regard to the size of class during this period was in the marked decrease in the number of classes having between 30 and 39 pupils. During the school year, 1918-19, the size of slightly more than one-third of the classes fell within these limits. During the other two school years, covered by this study, less than one-fifth of the classes came within these limits of size. This change is offset by a corresponding increase in the number of classes having between 40 and 49 pupils. Although the span of years covered by this table is insufficient to justify conclusions with reference to the trend of the size of class, Table I suggests that there is a tendency, in

TABLE I. SIZE OF CLASSES IN THE ELEMENTARY SCHOOLS OF 180 ILLINOIS CITIES FOR 1918-21.

Number in Classes	1918–19		1919–20		1920–21	
	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes	Number of Classes	Percent of Classes
50..........	313	3.3	342	3.5	350	3.4
40–49.......	5134	54.5	6865	70.0	7161	68.8
30–39.......	3264	34.6	1845	18.8	2050	19.7
20–29.......	598	6.3	646	6.6	719	6.9
Less than 20	113	1.3	110	1.1	123	1.2
Total.......	9422	100.0	9808	100.0	10403	100.00
Median.....		41.4		43.4		43.2

the elementary schools of Illinois outside of Chicago, to assign from 40 to 50 pupils to a teacher.

Size of class in Chicago public schools, October, 1920. Because the two investigations to be reported later were carried on, for the most part, in certain schools in Chicago, it was thought desirable to compile separately the facts relating to class size in Chicago. The information was taken from the records of the superintendent of schools. One-half of the elementary schools were selected at random, but all of the high schools were included. Information, with reference to the size of class in each grade, is summarized in Table II. Below the ninth grade, the classes are noticeably larger than in the high school. The median size of class is approximately 46 pupils. In the high school, a greater degree of variability is shown in the size of class. Approximately 14 per cent of the classes have fewer than 20 pupils. The median size of class for the high school is slightly over 30 pupils. It will be noted that the classes for the first year are larger than those for the following years. Of the 183 classes, 178 reported as having from 50 to 54 pupils, are classes in physical training. A number of these have more than 54 pupils.

The size of high school classes is given by subjects in Table III. The median size of class for the different subjects ranges from 20.0,

TABLE II. SIZE OF CLASSES IN CHICAGO SCHOOLS, OCTOBER, 1920.

Grades	Less than 10	Number in class									55 and over	Total	Median
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54			
1			3	1	6	20	40	120	248	87	36	561	46.8
2			1	2	1	6	11	102	247	38	1	409	46.7
3			1	1	1	3	18	96	211	48	3	382	46.7
4				1	2	3	15	90	207	39		257	46.6
5			1			2	4	79	215	48	1	350	47.1
6			3	3	1	2	12	103	183	27	4	338	46.2
7					2	2	15	92	180	32	3	326	46.4
8				1		4	54	79	117	15	48	318	44.8
Total Elementary Schools			9	9	13	42	169	761	1618	334	96	3041	46.6
9	36	108	224	556	613	850	683	254	98	92		3474	32.3
10	24	84	202	342	378	401	283	136	17	69		1936	29.3
11	14	60	90	178	148	189	119	31	10	19		858	27.8
12	15	27	64	67	81	93	77	29	2	3		458	28.4
Total H. S.	89	279	580	1143	1220	1533	1162	450	87	183		6726	30.2

for German, up to 34.5, for arithmetic. If physical education is included, the maximum median class is 43.8. However, the most significant aspect of the table is the wide variation in the size of class for a given subject. With few exceptions, there are, in each subject, classes having 10 or less pupils and also classes having more than 45 pupils. The exceptions are shop work, in which the number of pupils is probably limited by the equipment, office practise, of which there are only 37 classes in the entire school system, chemistry, home economics, botany, zoology and agriculture, in which equipment again probably limits the size of class. In German there are no classes having more than 34 pupils, but there are only 8 classes in the entire system.

Opinions of city superintendents in regard to the best size of class. A questionnaire was sent to the city superintendents of public schools in all cities in the United States having a population

TABLE III. SIZE OF CLASSES BY SUBJECTS IN CHICAGO HIGH SCHOOLS, OCTOBER, 1920.

Subject	Less than 10	Number of pupils in class									Total	Median
		10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54		
Botany, Zoology and Agriculture	6	12	24	46	49	51	22	5			215	27.0
French	6	14	33	51	35	38	15	8	2	1	203	24.7
Home Economics	7	27	42	57	54	18	11	3			219	22.9
Geography	1		17	17	19	14	19	8	1		96	28.4
Typewriting	1	1	12	36	50	70	74	29	11		284	33.0
Art	7	22	35	44	44	59	28	11	2		252	27.0
Physiology			1	13	17	36	44	12	1	1	125	34.4
German	2	1	1		3	1					8	20.0
Chemistry	2	5	15	21	19	39	3	1			105	27.5
Office Practise	3	4	2	11	4	6	7				37	24.3
Spanish	2	15	18	28	66	59	41	16	2		247	29.6
Physical Education	5	3	19	21	42	61	40	47	39	178	455	43.8
Latin	3	15	20	31	46	52	56	21	1		245	30.8
Mechanical drawing	22	29	63	136	97	67	31	9	1		455	24.2
Algebra	1	2	19	36	73	133	115	27	2		408	32.7
Geometry	1	9	16	36	62	79	51	13	3		270	30.6
Arithmetic			5	18	25	43	57	24	1		173	34.5
Bookkeeping	2	15	14	23	28	27	30	17	1		157	29.4
Stenography	1	9	23	55	53	72	65	17	4		299	30.6
Physics	1	2	12	29	18	40	22	5	1		130	30.4
Shop	8	66	94	228	101	16	3				516	22.0
History	4	13	27	49	37	92	86	63	3	1	375	33.1
English	3	15	60	143	228	352	278	85	10	2	1176	32.0
General Science	1		8	14	50	108	65	28	2		276	33.0

TABLE IV. IDEAL SIZE OF CLASSES AS INDICATED BY 270 SUPERINTENDENTS IN CITIES OF 25,000 OR MORE POPULATION.

Score	Grades			
	1, 2, 3,	4, 5, 6,	7, 8, 9,	10, 11, 12
55–59	1			
50–54		1	1	
45–49	2	2		1
40–44	16	18	15	2
35–39	53	73	35	4
30–34	95	95	104	33
25–29	52	42	58	95
20–24	20	21	48	95
0–19	30	18	9	16
Total......	269	270	270	246
Median....	31.7	32.8	31.0	25.6

of 25,000 or more, as shown by the directory issued by the Bureau of Education. The questionnaire asked the superintendents to indicate the ideal size of class in their opinion for grades one to three, four to six, seven to nine, and ten to twelve. Replies from 270 cities are summarized in Table IV. One of the most interesting things about this table is the wide range of opinion which it indicates. A considerable number of superintendents would have fewer than 20 pupils in each class. Other superintendents appear to consider classes of 40 or more ideal. One superintendent indicates that he would be satisfied with classes of 55 pupils in the primary grades. The ideal median size of class below the high school ranges from 31.0, for the junior high school, to 32.8, for the intermediate grades. In the senior high school, the ideal median of class is 25.6 pupils.

The problem of class size. Although Table IV is based upon replies from superintendents distributed over the United States, and the preceding tables refer to existing conditions in Illinois, we are probably justified in pointing out that a marked difference exists between theory and practise. The prevailing practise in the elementary schools of the state outside of Chicago centers around classes

having from 40 to 45 pupils. The median ideal size of class is only slightly above 30. Thus, it appears that the practical problem in which school superintendents are interested relates to a determination of the relative efficiency of classes enrolling from 25 to 35 pupils as compared with classes enrolling from 40 to 45 pupils.

For the high school, we have no data relating to the size of class in Illinois except for the city of Chicago. The median ideal size is approximately 25. The median actual size is approximately 30. It, therefore, appears that the practical problem in the high school relates to the relative efficiency of classes enrolling from 20 to 30 pupils as compared to those enrolling from 25 to 35.

CHAPTER III

RELATION OF SIZE OF CLASS IN ELEMENTARY SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the elementary school. In order to study the effect of variations in the size of class upon the achievements of pupils, it is necessary to hold constant or to measure the other factors which affect their achievements. It was planned to keep the teacher constant by having both a large class and a small class taught by the same teacher. Since the plan of organization in the elementary school makes it impossible for the same teacher to instruct two classes at the same time, it was necessary to have a teacher instruct the two types of classes during two consecutive semesters. It was arranged to have some of the teachers instruct a large class during the first semester and a small class during the second semester. Other teachers instructed a small class during the first semester and a large class during the second semester.[1]

In order to keep the pupil material as nearly constant as possible, "one hundred percent promotion" was secured at the end of the first semester in all of the experimental groups. When a teacher instructed a large class during the first semester a number of pupils were sent to another teacher at the beginning of the second semester. The pupils remaining formed a small class. In doing this, an effort was made to select pupils so that those remaining would form a small class having approximately the same average mental age and the same variability of this trait. When a teacher instructed a small class during the first semester, pupils were added at the beginning of the second semester, but care was exercised to have these pupils such that the average mental age of the class would not be materially affected.

This investigation, which began in October, 1920, was confined to classes in grades II, V, and VII. Some of the experimental groups were organized in the B sections of these grades and the others in the A sections. At the beginning of the second semester the B sec-

[1]This investigation was carried on in five elementary schools in Chicago: Washington, Cleveland, Lowell, Farragut, and Hibbard.

tions became A sections of the same grade and the A sections became B sections of the next higher grade. We shall, however, refer to the grades simply as II, V, and VII. Data for only those pupils who attended both the large class and the small class and who took all of the tests given are included in the following tabulations.

The size of the experimental classes. In the second grade there were eleven experimental classes, three small the first semester and large the second, and eight of the opposite type. If one class, enrolling only 18 pupils when considered a large class, is excluded, the small classes range from 33 to 44 and the large from 45 to 54. The differences in the size of the paired groups range from 4 to 13, the average being approximately eleven pupils. In the fifth grade there were thirteen experimental classes, three small the first semester and large the second, and ten of the opposite type. The small classes range in size from 33 to 45 and the large from 42 to 52. The differences in the size of the paired groups range from 4 to 14, the average difference being approximately 9. In the seventh grade there were only five experimental classes, three of one type and two of the other. The small classes ranged from 35 to 44 and the large from 42 to 49. The average difference in size was approximately 7. It should be noted that in both the fifth and seventh grades there is some overlapping in the size of the two types of classes. Some "large classes" are smaller than certain "small classes."

Data collected. In the second grade, the Dearborn Group Intelligence Test and Pressey Primer Scale were given at the beginning of the experiment. In this grade, achievement was measured by giving the Indiana Scale of Attainment, No. 1. Form 1 was given in October, Form 2 in January, and Form 1 was used again at the end of the year. In grades V and VII, the Illinois General Intelligence Scale was used. The achievements of the pupils were measured in arithmetic, silent reading, language, and spelling. In arithmetic and reading, the measurements were secured by means of the tests included in the Illinois Examination. Form 1 was used for the first and third testings and Form 2 for the second. In language, Charters' Diagnostic Language Test for pronouns was used. Form 1 was given in October and in May. Form 2 was used for the January testing. In spelling, 20 words were selected from columns N and R of the Buckingham Extension of the Ayres' Spelling Scale.

For the first and third testings, the words were selected by beginning at the bottom of these columns and choosing alternate words. The words for the second testing were taken from these columns, beginning with the next to the last word and taking alternate ones.

Administration of tests and collection of data. All the tests were administered and scored by the teacher. As a preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests and the plan of administration. In this connection, the tests were administered to the teachers in order to illustrate to them the procedure to be used with the pupils. All of the tests are highly objective with reference to the scoring, and samplings of the test papers failed to reveal any large errors in this work. The teachers reported the data for each pupil on an individual record card. This card contained spaces for each score and each test as well as for data with reference to the size of class in which the pupil was taught during each semester. The dates of testing were approximately as follow: October 20th, February 20th, and May 20th.

Method of summarizing the data. The data summarized were limited to the scores of only those pupils who were members of both a large class and a small class, and who were present at all three testing periods. The scores of all pupils in a grade (including both A and B sections), who had been taught in the same type of class, were assembled for each of the three testings. For example, the October arithmetic scores for all fifth grade pupils who were taught in large classes during the first semester and in small classes during the second semester were assembled in one distribution. Another distribution was made for the January scores and a third one for the May scores. Three corresponding distributions were made for the arithmetic scores of the pupils who were taught in small classes during the first semester and in large classes during the second. Thus, there were obtained six distributions for each achievement score. One measure of the gain in achievement made by a group of pupils during the first semester was found by subtracting the *average* of the October scores from the *average* of the January scores. Another measure of the gain was found by subtracting the *median* October score from the *median* January score. In a similar manner, the gains for the second semester were obtained by sub-

tracting the average and the median January scores from the corresponding May scores.

In calculating these gains, no account was taken of the possible non-equivalence of the different forms of the tests used. In fact, no accurate information concerning the equivalence of the duplicate forms is available, except for the tests in reading and arithmetic. The duplicate forms of these two tests have been shown to be approximately equal.[2] Since Form 1 was used twice, and the average and the median scores calculated from it were used both as subtrahends and minuends, any non-equivalence of the forms will not affect the comparisons of gains made in the following table.

The scores of the different tests are expressed in terms of different units. Thus, before any combination of the results from the different tests can be made, it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test and in one grade as in another. On the basis of this assumption, the stardard deviation was calculated for six of the different distributions of scores for each test in a given grade, and the average of these six standard deviations was used as a divisor to reduce the gains to a basis of a common unit. For example, during the first semester the fifth grade pupils taught in large classes in arithmetic made a gain of 12.0 points.[3] During the second semester they made a gain of 7.2 points. The gains for the pupils taught in small classes in arithmetic were 6.5 and 5.65. The average standard deviation of the six distributions of arithmetic scores is 17.441. Dividing these gains by this average standard deviation, we secure as quotients the entries (.68, .41, .36, and .32) to be found in Table V.

In calculating the average gains for the two types of classes, the simple average of the two gains has been used rather than the weighted average, although the two groups are not even approximately equivalent in size. Since our purpose in taking this average is to eliminate any differences in the course of study or in the edu-

---

[2]Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin, No. 6. Urbana: University of Illinois, 1921.

[3]These gains were calculated from the average of the scores.

TABLE V. THE GAINS MADE BY PUPILS IN LARGE CLASSES AND IN SMALL CLASSES. CLASSES OF ALL SIZES INCLUDED

	No. of pupils	Average size of class		Arithmetic Gains		Dif.	Reading Comprehension Gains		Dif.	Reading Rate Gains		Dif.	Language Gains		Dif.	Spelling Gains		Dif.	Totals		Dif.
		Large	Small	Large	Small		Large	Small		Large	Small		Large	Small		Large	Small		Large	Small	
GRADE II.																					
Average Scores:																					
Large 1st S.	244	44.8	33.1	1.89	.59		1.17	.59		1.27	.38					1.65	.88		5.98	2.44	
Small 1st S.	109	48.0	38.0	.30	1.48		.53	1.09		.51	1.00					−.13	1.83		1.21	5.40	
Average..........				1.09	1.04	.05	.85	.84	.01	.89	.69	.20				.76	1.36	−.60	3.59	3.93	−.34
Median Scores:																					
Large 1st S.	244	44.8	33.1	2.25	.70		1.28	.80		1.09	.60					2.06	.95		6.68	3.05	
Small 1st S.	109	48.0	38.0	.33	1.94		.80	1.05		.53	.97					.50	1.64		2.16	5.60	
Average..........				1.29	1.32	−.03	1.04	.93	.11	.81	.78	.03				1.28	1.29	−.01	4.42	4.32	.10
GRADE V																					
Average Scores:																					
Large 1st S.	325	45.8	35.7	.68	.41		.20	.88		.09	.72		.89	1.22		.02	.62		1.88	3.85	
Small 1st S.	132	51.0	44.0	.36	.32		.55	.38		.76	.28		.93	.96		.38	.29		2.98	2.23	
Average..........				.52	.37	.15	.38	.63	−.25	.43	.50	−.07	.91	1.08	−.17	.20	.46	−.26	2.44	3.04	−.60
Median Scores:																					
Large 1st S.	325	45.8	35.7	.78	.36		.32	1.12		.08	1.14		.79	1.77		.01	.67		1.98	5.06	
Small 1st S.	132	51.0	44.0	.21	.43		.75	.29		.93	−.03		.98	.99		.43	.26		3.30	1.94	
Average..........				.50	.40	.10	.54	.71	−.17	.51	.56	−.05	.89	1.38	−.49	.22	.46	−.24	2.66	3.51	−.85
GRADE VII.																					
Average Scores:																					
Large 1st S.	104	45.0	36.0	.85	1.29		.69	.72		.74	.45		.57	1.26		.92	.21		3.77	3.93	
Small 1st S.	73	47.0	42.0	.53	.47		.42	.19		.64	−.61		.62	1.00		.15	.44		2.36	1.49	
Average..........				.69	.88	−.19	.56	.46	.10	.69	−.08	.77	.59	1.13	−.54	.54	.33	.21	3.07	2.72	.35
Median Scores:																					
Large 1st S.	104	45.0	36.0	.92	1.28		.76	.86		.67	1.01		.50	1.34		.98	.47		3.83	4.96	
Small 1st S.	73	47.0	42.0	.65	.66		.45	.21		.56	−.45		1.09	.99		.24	.64		2.99	2.05	
Average..........				.79	.97	−.18	.61	.54	.07	.62	.28	.34	.79	1.17	−.38	.61	.56	.05	3.42	3.52	−.10
Totals of averages of Average Scores...				2.30	2.29	.01	1.79	.193	−.14	2.01	1.11	.90	1.50	2.21	−.71	1.50	2.15	−.65	9.10	9.69	−.59
Totals of averages of Median Scores....				2.58	2.69	−.11	2.19	2.18	.01	1.94	1.62	.32	1.68	2.55	−.87	2.11	2.31	−.20	10.50	11.35	−.85

cational opportunities offered in the two semesters and also any practise effect due to acquaintance with the tests, it seemed unwise to weight the averages on the basis of the number of pupils in the two groups. To have used the weighted averages in this case would have resulted in giving greater weight to the gains made by one group of pupils simply because this group happened to be larger.

Achievements of the two groups approximately equal. Table V[4] summarizes the data with reference to the gains made by the two groups. The column headed "Number of pupils" gives the number of pupils whose records were used in the tabulations. (A pupil's record was discarded if he was not a member of both the small class and the large class and if he did not take all tests.) The average size of class is computed from the total enrollment. The computation of the gains has just been explained. In interpreting the table, attention should be focused upon the differences. A positive difference means that the large class is superior in achievement, and a negative difference that the small class is superior. At the bottom of each difference column the differences, calculated from the averages and also from the medians, are summarized. This summary is a total and not an average. To find the average it is necessary to divide by 3. All of the totals of the differences fall between +1.00 and —1.00. Six of the 10 differences are negative, and only in the case of reading rate are the difference between the averages and the difference between the medians both positive. In the last three columns of the table, we have the totals and not averages. To find the average, it would be necessary to divide by 4 in the second grade and by 5 in each of the other two grades. Here, again, the negative differences predominate, although none of them are very large. The last two entries in the last column are essentially grand totals and may be considered to summarize the entire table. To find the average difference, each of these numbers should be divided by 14. The quotients obtained would be —.04 and —.06. Thus, in general, this table indicates that there is little if any superiority in the achievements of pupils in the small classes over those of pupils in the larger classes.

---

[4]The entries in the column headed "Reading Rate" in the second grade are based upon the Pressey Word Recognition Test.

An examination of Table V reveals the fact that, in general, the gains between the first and second testings are much larger than those between the second and third testings. This condition emphasizes the necessity for equalizing the effect of acquaintance with the test and practise effect. If all of the experimental groups had been taught as large classes the first semester and small classes the second, the gains for the large classes would greatly exceed the gains for the small classes; but this would be due primarily to the effect of acquaintance with the test and the practise effect.

When Table V is examined with reference to the conditions in the different grades we find that the gains are relatively greater for the small classes in the fifth grade than in either the second or the seventh grade. However, the number of pupils is so small for the two groups in the seventh grade and the groups differ so little in size that only slight significance can be attached to the results. Even in the second and fifth grades it is unfortunate that the experimental groups are not more nearly equal in size. It is possible that if the experiment had included a larger number of classes which were small the first semester and large the second different results might have been obtained.

When the gains for the different subjects are examined we find that in language the gains for the small classes are consistently greater than the gains for the large classes. This is not true for any other subject. Although in spelling the total of the gains is distinctly negative, in both arithmetic and reading comprehension the total when computed by one method is negative and in the other case is approximately 0. Reading rate in the seventh grade is the only case in which the large class is distinctly superior in achievement.

Conclusion: the relation of the size of class to school efficiency. Since Table V indicates that, on the whole, there is little difference between the achievements of the pupils when taught in large classes and their achievements when taught in small classes, one might infer that the efficiency of a school would be materially increased by the formation of large classes, because the educational output would be approximately the same and the educational investment would be materially decreased. However, it is doubtful that the present investigation justifies such a conclusion. In the first place, it is obvious that only certain achievements of pupils have

been measured. Even in the fields of the four subjects in which tests were given we are not justified in claiming that all achievements of the pupils were measured. The arithmetic tests used were confined to the operations and to only certain types of examples within this division of arithmetic. In silent reading, the test used is very limited in scope. Similarly, the tests in language and spelling possess very definite limitations with respect to scope. There is some justification for assuming that the measurements made may be considered indices of the total achievements of the pupils not only in the fields of the four subjects in which the tests were given but also in the field of instruction in the grades concerned. However, the thesis that the measures of achievement secured in this investigation are indices of the total achievement is largely an assumption, and in interpreting the results it is necessary to recognize this fact. It is possible that, if other tests had been used or if the achievements of the pupils had been more completely measured by including tests in other subjects, the results might have been different.

In the second place, it must be remembered that the size of the "small classes" was not less than 33 (with one exception), and in a few cases the enrollment was as much as 44 or 45. The large classes ranged in size from 42 to 54. The average difference between the pairs of experimental groups ranged from 7 in the seventh grade to 11 in the second grade. These conditions with reference to the size of the experimental groups constitute a very significant limitation of the investigation. One is not warranted in making inferences from the facts of Table V with reference to the relative efficiency of classes of 20 to 25 pupils as compared with classes of 35 to 45 pupils. No application should be made except within the limits of size defined by the experimental groups.

CHAPTER IV

RELATION OF SIZE OF CLASS IN HIGH SCHOOL TO SCHOOL EFFICIENCY

General plan of the study in the high school. In the investigation in the high school it was arranged to have both a large class and a small class in a given subject instructed by the same teacher during the same semester. This made it necessary to restrict the experiment to teachers who were instructing two or more sections of the same subject. When a teacher was instructing two sections, pupils were shifted on the basis of their intelligence scores so that the average quality of pupil material was approximately the same in the two sections. Thus, both classes would include some bright, some medium, and some dull pupils. When a teacher had four sections of the same subject, the pupils were shifted so that a large class and a small class would be obtained, consisting of relatively bright pupils. The less capable pupils were also divided into two classes, one large and one small.

In establishing the two types of class, there was considerable variation in the size of both the large classes and the small classes. The small classes varied in size from 12 to 26 pupils. The large classes varied in size from 23 to 45 pupils. The average size of the large classes was 36.5 and that of the small classes, 20.8. The differences in the size of the paired classes ranged from 6 to 26.

Source of data. This study was carried on in four large high schools in Chicago and in the high schools of three other Illinois cities.[1] During the first semester of 1920-21, the experiment was carried on in beginning tenth grade classes. During the second semester of that year the study was confined to classes completing the first year of high school work. In the following tables no distinction is made between classes for the two grades. Records were secured for 67 pairs of classes, enrolling 3,821 pupils. The project

[1]The high schools in Chicago were Lane Technical, Tilden Technical, Harrison Technical, and Hyde Park. The three other Illinois cities were Macomb, Shelbyville, and West Aurora.

was begun by giving the Terman Group Test of Mental Ability to all pupils concerned, about October 15, 1920. Some pupils in the high schools outside of Chicago were given the Illinois General Intelligence Scale or the Chicago Group Intelligence Test. As soon as the results of the testing could be assembled, the large classes and small classes were arbitrarily formed, following the method indicated above.

The educational output, or the achievements of the pupils, was measured by requiring each teacher to give the same final examination to both types of classes. A check upon this measurement of achievement was secured by using the "term grades" of the pupils. It is generally recognized that "term grades," as well as examinations set by teachers, are highly subjective. However, in this case the same teacher administered the same examination to both the small class and the large class. The same teacher also gave the "term grades." Thus, there is in no place a comparison between either "term grades" or "examination grades" given by different teachers. This tends to eliminate the subjective factor of these measures. In addition, it may be noted that we are concerned with the average "grades" of relatively large groups of pupils and not with the "grades" of individual pupils.

A limitation. The plan of carrying on the experiment implies the assumption that the achievements of the pupils in the large classes were equal to the achievements of the pupils in the small classes at the beginning of the experimental period. There was no attempt to measure the achievements of pupils in the subjects concerned at the beginning of the experiment. In the first semester, the two types of classes were not organized until after the high schools had been in session several weeks. In the case of the classes used during the second semester, the pupils had received an entire semester of instruction in regular classes. It is true that the sections were formed so that the average general intelligence of the paired classes was approximately equal, but this probably does not justify the assumption of equivalent achievements.

Details of administration. The intelligence tests were administered and scored by the teachers. As preparation for this work, the teachers were called together and given definite instructions concerning the nature of the tests. The tests were also administered

to them. They were then required to score their own papers. All the tests used were highly objective with reference to scoring, and a sampling of the test papers of the pupils failed to reveal any large errors in this work. Since, in the use of the intelligence test scores, comparisons are always made between the pupils or groups of pupils under the same teacher, variations in the administration of the tests, due to differences between teachers, would not be significant.

At the close of each semester the teachers were asked to report both the "final grade" and the "examination grade" for each student.[3] In most cases, the "examination grades" were reported in terms of percents. The "final grades" were generally reported in terms of letters, as follow:

S—Superior
E—Excellent
G—Good
F—Fair
D—Failure

For the purpose of combining "grades," these letters were assumed to represent the following numerical ranges on a scale of 100 percent:

S is equivalent to 95 and over
E is equivalent to 85 to 94
G is equivalent to 75 to 84
F is equivalent to 65 to 74
D is equivalent to 55 to 64[4]

Plan of summarizing data. The data collected were summarized to show the differences, if any, which existed between the final achievements of pupils in the large classes and of pupils in the small classes. Two methods of summarizing were employed. In the first, the achievements of all pupils were considered. According to the second method, the records considered were limited to those of pupils in the large classes who could be paired with pupils having identical scores on the intelligence test in the corresponding small

[3]The following tables, in which the data for high school classes are summarized, indicate that the "examination grades" were not received from certain classes.

[4]The midpoints of these intervals were presumably used as the numerical equivalents of the grades expressed in terms of letter. However, Mr. Stevenson's report yields no information on this point.

classes. For both of these methods two tabulations have been made. The first includes all classes, and the second only those pairs of classes in which the large class was at least twice the size of the small class.

Differences in achievements when all pupils are considered. Table VI illustrates the first method of summarizing the data for 22 pairs of English classes. In the second and third columns of this table the enrollment in the large classes and in the small classes is given. The quantities recorded in the three columns headed "Difference" are found by subtracting the quantities in the two columns immediately preceding the difference column. The number for the small class is, in every case, taken from that for the large. Therefore, a positive difference means that the large class is superior in the trait concerned, and a negative difference, that the small class is superior. The line at the bottom of the table gives the average for each column. These averages may be taken as summarizing the data collected from these 22 pairs of English classes, although the items combined are not entirely comparable. For example, different general intelligence tests were used in different classes. The present writer has not been able to ascertain the particular intelligence test given to any pair of these 22 pairs of classes. It is, however, difficult to explain the extreme differences between the average intelligence scores of classes 10 and 11 on any basis other than the use of different tests in these two pairs of classes. Furthermore, it is not unlikely that different passing marks are in use in the different schools in which the pairs of classes were taught. If this is the case, in the case of different pairs of classes, both the average term grades and the average examination grades are on different scales.

It should be noted that, although an effort was made to organize a large class and a small class so that the average intelligence scores would be approximately the same for the two classes, this was not always accomplished. Because of conflicts or other conditions that could not be disregarded, it was not always possible to shift pupils from one section to the other so as to set up the desired class organization.

Table VII summarizes the averages for the classes in the different subjects. In interpreting this table, it is necessary to bear in mind that, with the exception of English and algebra, the number of

TABLE VI. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND EXAMINATION GRADES OF 22 LARGE AND 22 SMALL ENGLISH CLASSES.

Teacher	Size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
	Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
1	36	12	58.7	57.2	1.5	85.8	77.5	8.3	85.3	77.5	7.8
2	37	17	95.2	99.7	−4.5	69.4	71.2	−1.8			
3	34	17	54.0	50.7	3.3	76.8	79.4	−2.6			
4	40	18	111.5	114.2	−2.7	82.1	81.2	0.9	77.8	85.1	−7.3
5	44	18	92.3	88.4	3.9	72.8	72.2	0.6	74.1	73.6	0.5
6	38	19	98.1	100.2	−2.1	74.2	80.6	−6.4	71.6	80.0	−8.4
7	35	19	138.7	137.1	1.6	77.1	81.6	−4.5	79.8	83.2	−3.4
8	32	19	94.9	87.8	7.1	72.4	75.6	−3.2	69.2	74.1	−4.9
9	41	20	77.4	72.7	4.7	75.8	77.0	−1.2	75.8	75.3	0.5
10	38	20	90.1	89.7	0.4	75.8	77.0	−1.2	80.0	78.0	2.0
11	35	21	55.8	55.7	0.1	70.9	73.0	−2.1	71.2	75.8	−4.6
12	43	21	136.1	142.0	−5.9	79.8	82.7	−2.9	78.3	81.1	−2.8
13	41	21	116.1	110.0	6.1	75.0	73.3	1.7	75.3	72.4	2.9
14	38	21	120.3	122.1	−1.8	87.5	86.9	0.6	88.3	86.8	1.5
15	36	21	97.4	98.4	−1.0	75.2	75.9	−0.7			
16	39	21	95.5	95.6	−0.1	76.4	82.9	−6.5	77.4	83.0	−5.6
17	41	22	131.2	119.9	11.3	76.8	77.8	−1.0	78.4	74.6	3.8
18	40	22	101.0	89.6	11.4	80.4	80.0	0.4	76.7	80.2	−3.5
19	32	24	127.1	143.7	−16.6	76.7	75.8	0.9	81.5	81.0	0.5
20	37	24	130.3	127.1	3.2	79.7	78.6	1.1	80.7	82.9	−2.2
21	38	26	105.7	99.3	6.4	78.3	82.3	−4.0	80.3	83.8	−3.5
22	38	26	130.4	135.2	−4.8	70.5	76.5	−6.0	52.2	69.0	−16.8
Total......	833	449	2257.8	2236.0	21.8	1689.4	1719.0	29.6	1453.9	1497.4	−43.5
Average....	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3

pairs of classes is so small that we probably should not consider the result representative. Certainly, little if any significance can be attached to the results for Latin, history, and French. The average for the 67 pairs of classes is the weighted average, so that one pair of classes does not influence this average any more than any other pair.

Table VIII presents a summary for those pairs of classes in which the large class is at least twice the size of the small class. By doing this, we are able to examine the achievements of pupils in pairs of classes where the difference in size is marked. The number of classes in any one school subject is so small, with the exception of English, that we are probably not justified in drawing any conclusions for the separate subjects.

Differences in the achievements of paired pupils. Table IX is similar to Table VII, the only difference being that it is based upon the records of only those pupils in the large classes who could be paired with pupils in the corresponding small classes, having the same scores on the general intelligence test. Since the pupils were paired on the basis of their intelligence scores, the average general intelligence of those taken from the large classes would be identical with the average general intelligence of those taken from the corresponding small classes. Hence, the average general intelligence scores are omitted. Table X is similar to Table VIII.

Interpretation of the tables. When all 67 pairs of classes are considered, the average of the general intelligence scores for the large classes is almost identical with that for the small classes, the difference being only one-tenth of a unit. This unit corresponds approximately to one month of mental age. We may, therefore, consider the pupils in the large classes equal in general intelligence to the pupils in the small classes. Both the average "term grade" and the average "examination grade" are slightly larger for the small classes.

The question of the significance of the difference of two averages is involved here. Both examination grades and final grades are known to be highly subjective and to involve a large error which is a combination of a constant error and a variable error. Variable errors tend to offset each other in an average because some of them are negative and some positive. On the other hand, constant errors

TABLE VII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF THE DIFFERENT SUBJECTS FOR 67 LARGE AND 67 SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	22	37.9	20.4	102.6	101.6	1.0	76.8	78.1	−1.3	76.5	78.8	−2.3
Algebra	18	35.7	20.3	99.6	102.2	−2.6	74.3	76.5	−2.2	76.3	77.2	−0.9
General Science	6	35.7	21.5	110.5	109.6	0.9	75.3	77.2	−1.9	74.8	76.1	−1.3
Botany	4	28.3	18.0	126.3	122.2	4.1	73.8	74.2	−0.4	71.9	72.9	−1.0
Stenography	3	36.7	22.0	83.6	85.9	−2.3	77.6	78.1	−0.5	78.4	80.5	−2.1
Geometry	3	40.7	22.7	128.1	121.1	7.0	75.8	76.6	−0.8	62.9	67.8	−4.9
Arithmetic	3	38.7	22.7	90.2	89.8	0.4	74.7	74.5	0.2	63.3	64.3	−1.0
Typewriting	3	39.3	24.7	83.4	87.9	−4.5	73.2	76.6	−3.4	80.1	83.0	−2.9
French	2	35.0	19.0	110.0	113.2	−3.2	73.7	70.9	2.8	71.2	71.4	−0.2
History	2	33.5	21.0	139.1	137.4	1.7	75.7	74.6	1.1	74.6	74.3	0.3
Latin	1	39.0	25.0	103.7	92.5	11.2	77.3	73.9	3.4	75.4	70.9	4.5
Average		36.5	20.8	104.1	104.0	0.1	75.4	76.7	−1.3	74.8	76.4	−1.6
Total	67											

TABLE VIII. COMPARISON OF AVERAGE INTELLIGENCE SCORES, AVERAGE TERM GRADES, AND AVERAGE EXAMINATION GRADES OF LARGE AND SMALL CLASSES WHERE THE LARGE CLASSES ARE AT LEAST TWICE THE SIZE OF THE SMALL CLASSES.

Subject	No. of pairs of classes	Average size		Average intelligence score		Difference	Average term grade		Difference	Average examination grade		Difference
		Large class	Small class	Large class	Small class		Large class	Small class		Large class	Small class	
English	8	39.1	17.8	90.4	90.6	−0.2	77.1	77.7	−0.6	77.2	78.8	−1.6
Algebra	5	39.6	18.2	96.2	100.4	−4.2	74.6	78.5	−3.9	85.6	86.3	−0.7
General Science	1	42.0	17.0	73.4	78.7	−5.3	70.4	70.0	0.4	70.4	70.0	0.4
Botany	1	33.0	15.0	128.8	117.6	11.2	78.1	71.8	6.3	78.4	75.5	2.9
Stenography	1	36.0	18.0	52.4	55.7	−3.3	73.3	72.9	0.4	71.3	76.5	−5.2
Average		38.9	17.7	91.2	92.4	−1.2	75.7	76.8	−1.1	79.1	80.2	−1.1
Total	16											

TABLE IX. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN 67 LARGE AND 67 SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	22	336	37.9	20.4	76.4	78.1	−1.7	76.6	78.8	−2.2
Algebra	18	298	35.7	20.3	74.3	76.3	−2.0	76.4	77.1	−0.7
General Science	6	102	35.7	21.5	76.2	78.7	−2.5	75.5	76.6	−1.1
Botany	4	36	28.3	18.0	73.5	72.9	0.6	68.8	70.5	−1.7
Arithmetic	3	56	38.7	22.7	73.8	74.3	−0.5	62.3	64.7	−2.4
Geometry	3	50	40.7	22.7	76.1	77.0	−0.9	61.9	68.2	−6.3
Stenography	3	44	36.7	22.0	76.3	79.8	−3.5	76.5	81.6	−5.1
Typewriting	3	42	39.3	24.7	73.7	76.2	−2.5	77.9	83.6	−5.7
French	2	25	35.0	19.0	71.2	70.0	1.2	68.9	70.0	−1.1
History	2	31	33.5	21.0	78.1	75.4	2.7	75.1	75.4	−0.3
Latin	1	14	39.0	25.0	73.6	79.3	−5.7	73.6	75.0	−1.4
Average			36.5	20.8	75.2	76.8	−1.6	74.4	76.5	−2.1
Total	67	1034								

TABLE X. COMPARISON OF AVERAGE TERM GRADES AND AVERAGE EXAMINATION GRADES OF PUPILS PAIRED ACCORDING TO EQUAL INTELLIGENCE SCORES IN LARGE AND SMALL CLASSES IN WHICH THE LARGE CLASSES ARE TWICE THE SIZE OF THE SMALL CLASSES.

Subject	Number of pairs of classes	Number of pairs of pupils	Average size		Average term grade of paired pupils		Difference	Average examination grade of paired pupils		Difference
			Large class	Small class	Large class	Small class		Large class	Small class	
English	8	116	39.1	17.8	76.6	77.7	−1.1	76.7	78.5	−1.8
Algebra	5	84	39.6	18.2	73.7	77.5	−3.8	83.6	86.3	−2.7
General Science	1	16	42.0	17.0	71.9	71.9	0.0	71.9	71.9	0.0
Stenography	1	11	36.0	18.0	72.7	72.7	0.0	70.6	77.6	−6.0
Botany	1	11	33.0	15.0	74.6	71.8	2.8	71.8	75.5	−3.7
Average			38.9	17.7	75.0	76.6	−1.6	77.9	80.3	−2.4
Total	16	238								

are not eliminated in averages. The number of pupils included in the 67 pairs of classes is sufficiently large so that the average grades include a variable error which is probably so small as to be negligible. For example, if we assume that the probable variable error of a grade is as much as 10 points, which is probably in excess of the actual probable variable error, the probable variable error of the average for the small classes would be less than three-tenths of one point. In the case of the large class, it would be materially less than in the small class. The constant errors, which are expressed in the tendency of some teachers to give higher grades than others, are probably included in both groups in approximately the same proportion, since the same teacher assigned grades to both a small class and a large class. If this is true, a difference of 1.3 between the average term grades is probably significant, although there is a reasonable doubt. This doubt is materially increased and probably the difference loses its significance when we recall that the achievements of the two groups of students were not measured at the beginning of the experimental period.

When we consider the summaries for the different subjects, we find the differences between the average achievements of the pupils in the two types of classes materially larger in a number of cases than the differences between the averages for the 67 pairs of classes. However, in interpreting these differences it is necessary to remember that the number of pupils is materially less and, hence, a difference must be materially greater in order to be significant. It is perhaps significant that negative differences predominate. This suggests that the average achievements in the small classes are somewhat superior to those in the large classes, but in drawing conclusions from this condition it is necessary to bear in mind the fact that the achievements of the students were not measured at the beginning of the experimental period.

When we turn to Table VIII, which is restricted to those classes in which the large class is at least twice the size of the small class, we find that the pupils in the small classes were slightly superior in general intelligence. If the size of class is a potent factor in determining the achievements of pupils, we should naturally expect to find a greater difference in the achievements of the two types of classes than we found in Table VIII. The fact that the difference

is less, both absolutely and relatively, suggests that the size of class is not a potent factor in determining the achievements of pupils. However, the number of pairs on which Table VIII is based is so small that no great importance should be attached to this observation.

When we examine Table IX, we find that the differences between the averages for the 67 pairs of classes are only slightly larger than those given in Table VII. In general, there are few significant differences to be noted in a comparison of these two tables. One of the most significant is the reversal of the relative achievements of the large class and the small class in Latin. In Table VII, when all pupils were considered, those taught in the large class were shown to be distinctly superior in achievement. When only the paired pupils were considered, those taught in the small class were distinctly superior in achievement. Considering the table as a whole, we are justified in asserting that it tends to corroborate the interpretations suggested for Table VII.[4]

Conclusion: relation of size of class in high school to school efficiency. The tables of this chapter show that at the end of the experimental period the achievements of the students in the two types of classes were approximately equal, and there is a slight indication that those taught in small classes were superior. Since the educational investment can be materially decreased by increasing the size of class in the high school, one might infer that the efficiency of the school would be increased by organizing classes enrolling from 35 to 40 students instead of classes enrolling from 20 to 25. In addition to the fact that there are several uncontrolled factors whose influence is unknown, it is necessary to bear in mind the exact conditions of the experiment. Since the same teachers taught both

[4]In his report, Mr. Stevenson attempted the further analysis of the data by ascertaining the percent of pairs of pupils in which the pupil in the small class received a higher grade than the corresponding pupil in the large class. When all pairs of pupils are considered, he shows that only in 40 percent of the pairs does the pupil in the small class surpass his mate in the large class. However, it is impossible to draw any conclusion from this fact, because we do not know the percent of pairs in which the two pupils received the same mark. Since, in a majority of the cases, the grades were reported in terms of letters and only five marks were recognized, it is reasonable to expect that in a relatively large percent of the cases both of the paired pupils received the same final grade.

a small class and a large class, there was no difference between the total amount of work done by the teachers who handled the large classes and the teachers who handled the small classes. In fact, they were the same teachers. Thus, this experiment failed to set up the conditions of large classes as a general plan of organization of a high school. It did, however, realize the conditions which not infrequently exist in the smaller high schools where it is desirable to have a few large classes assigned to teachers who are given compensating small classes or who have the number of classes reduced accordingly. The results of the experiment, therefore, can be applied only to those situations in which the teaching load is kept constant. In such cases the evidence collected indicates that approximately the same average achievement can be expected from the pupils taught in large classes as from those taught in small classes. In other words, the results of this experiment indicate that there is no loss of efficiency caused by organizing a few large classes if the other work assigned to the teacher is such that the teaching load is not increased.

One should recognize that the results of this experiment should not be applied to the question of the size of class where increasing the size of class results in a distinct increase in the teaching load. The instruction which students receive is given partly in the class room and partly through written work and individual conferences. In such subjects as English composition, algebra, and science requiring laboratory work, it is customary with most teachers to require a large amount of written work. A teacher who gives instruction to five classes of 40 students each has a much heavier teaching load than the teacher who instructs five classes of 20 students each, unless he introduces compensating changes in the amount of written work, in the method of handling it, and in the number of individual conferences. In such cases the question of class size is so intimately connected with the method of instruction that we are not justified in drawing any inferences from an investigation in which the method of instruction was assumed to be the same for both types of classes.

CHAPTER V

SUGGESTIONS FOR EDUCATIONAL EXPERIMENTATION[1]

The two studies described in Chapters III and IV make very slight contributions to a scientific determination of the relation between the size of class and the efficiency of a school system. They do, however, seem to the writer of this chapter suggestive with reference to the procedure of educational experimentation. The causes of the failure of these studies to produce reliable and significant results have been mentioned in the two preceding chapters but they may be summarized under two general heads: *first,* failure to set up and maintain appropriate experimental conditions and, *second,* the lack of adequate instruments for measuring the achievements of pupils.

A prerequisite for setting up and maintaining appropriate experimental conditions is a complete analysis of the problem being studied. The various factors involved must be recognized by the experimenter and the possibility of any relations which may exist between these factors must also be considered. For example, in the high school many factors contribute to the achievements of students, or the educational output of the school. In addition to the size of class, which is the factor whose relationship to school efficiency was studied, it is necessary to recognize methods of instruction, the personality and enthusiasm of the teacher, the discipline of the class and of the school, the general spirit of the school, the general attitude of the community toward the school, the time of day when the class recites, the textbooks used, the equipment, including the building, the "spiritual composition" of the class, the general intelligence of the students, their nationality, their past experience, both in school and out of school, the stage of advancement reached in their education, and possibly even other factors. It appears likely that certain of these factors are interrelated. The size of class is likely to affect the enthusiasm of the teacher, particularly if it determines the in-

[1]This chapter has no counterpart in the report submitted by Mr. Stevenson. It is entirely the contribution of the present director of the Bureau of Educational Research.

structional load carried by the teacher. It also appears that some relationship exists between the size of class and methods of instruction, and between the size of class and discipline. The existence of a functional relationship between two or more factors makes it impossible under normal conditions to produce variations in one factor without, at the same time, causing changes in the others. Failure to analyze the problem sufficiently will frequently cause the results of an educational experimentation to have little significance, and consequently the time and money invested in the study will be largely wasted.

When one considers the total product of the educational process one cannot fail to become impressed with the inadequacy of our present educational tests as instruments for the measurement of the various elements of this product. In the study relating to the size of class in the elementary school a more elaborate group of tests might have been used, but even if this had been done, it does not appear likely that one would be justified in asserting that the total product of education had been measured. In the high school no standardized educational tests were used. It was attempted to have the achievements measured by means of a final examination and by the term "grade" given to the students. The writer of this chapter is not aware of the considerations which resulted in the decision not to use any of the standardized educational tests that have been devised in the field of high school subjects, but it is likely that this decision was due to the belief that none of the available educational tests were sufficiently satisfactory measuring instruments to justify their use in this investigation. The present writer is inclined to share this belief. Thus, we cannot escape the conclusion that at the present time we do not have available instruments for measuring the outcomes of teaching which permit reliable educational experimentation when it is necessary to measure the total product of instruction.

Incidentally, attention may be called to the fact that more consideration should be given to the errors involved in the data and to the effect of these errors upon the results of statistical calculations. For example, it is highly important to know what significance should be attached to a difference between two averages.

In view of the difficulties encountered in setting up and maintaining appropriate experimental conditions and in view of the im-

perfections and limitations of our present educational tests, it is not inappropriate to question the wisdom of undertaking such complex educational experimentation as has been described in this monograph. It is true that there are many educational problems which are highly important. For example, an increase in the size of class would result in a material reduction in the educational expenditures for instruction. A reliable scientific determination of the relationship existing between the size of class and school efficiency would be a valuable contribution, but it is doubtful whether such a determination of this relation is at the present time possible.

To the present writer, it appears highly important that those engaged in educational research should give very careful consideration to the sort of problems to which they devote their energies. It is, of course, necessary that pioneer work be done, and in studies of this type, it is not always possible to anticipate the limitations of one's procedure. As a result it may become necessary to "scrap" a project because of the defects that appear in the course of one's work. Such losses are unavoidable in extending the frontiers of educational research. When an investigation is not pioneer work an experimenter should determine the limitations of his procedure in advance, and unless it appears likely that information of considerable value will be secured in spite of the limitations the investigation should not be undertaken. The fact that a problem is important does not justify its study. Educational experimentation which involves the use of faulty method and faulty instruments not only fails to make adequate contributions to our educational progress, but, more important, it tends to reflect unfavorably upon the application of the methods of research to the field of education.

BULLETIN NO. 11

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

RELATION OF SECTIONING A CLASS TO THE EFFECTIVENESS OF INSTRUCTION

by

WALTER S. MONROE.

PRICE 15 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

PREFACE

The educational experiment reported in this bulletin was initiated by the former Director of the Bureau of Educational Research and the data collected under his supervision. The present Director of the Bureau is responsible for the tabulation of the data and for the preparation of this report.

This investigation was made possible through the cooperation of Superintendent Peter A. Mortenson and of certain principals and teachers of the Chicago Public Schools. Not only did they cooperate in the collection of the data but they also made substantial contributions to the project by supplying test materials. The writer is glad to acknowledge the indebtedness of the Bureau of Educational Research to all who contributed to this project.

WALTER S. MONROE, *Director*

November 10, 1922

Relation of Sectioning a Class to the Effectiveness of Instruction

The problem. The purpose of this educational experiment was to determine the relative effect upon the achievements in certain school subjects of three plans of sectioning a class. A "class" is defined as the total number of children assigned to a teacher for instruction even though they may be divided into two or more groups for instructional purposes. The three plans of sectioning a class considered in this investigation are: (1) teaching a class as a single unit; (2) dividing the class into two equal groups approximately equivalent with respect to general intelligence; (3) dividing the class into three equal groups approximately equivalent with respect to general intelligence. When a class is taught as one group, all of the pupils recite at the same time. Following the recitation there is a period for study. Thus under this plan the work of the teacher alternates between "hearing classes" and supervising the study of the pupils. When a class is taught as two sections, one group recites while the other group studies. In this case the teacher's time is almost wholly devoted to "hearing classes." Any supervision of the study of the pupils is of necessity given incidentally and at irregular intervals when the teacher is fortunate enough to have a few minutes of leisure during a recitation period. When a class is divided into three sections, the conditions are much the same except that necessarily the length of the recitation periods is reduced. In general pupils of one section study during the recitation periods of the other two sections.

The specific problem of this investigation was to determine the relative effect of these three plans of sectioning a class upon the direct results of instruction in certain school subjects. In other words this investigation sought to answer the question, "Which is the best plan of sectioning a class?"

General plan of the experiment. If it were possible to secure three groups of classes so that all factors which affect the results of instruction were equivalent in the beginning of the experiment and could be controlled throughout the experimental period, the simplest procedure would be to have one group of classes taught as a unit, another group taught in two sections and a third group in three sec-

tions. However, it would be difficult, if not impossible, to secure exact equivalence of teaching ability and of pupil material. Our facilities for measuring the ability of teachers are extremely crude and at best it would be difficult to demonstrate that any differences found in the results of instruction were not produced largely by differences in teaching ability. It is true that we have a number of general intelligence tests which might be used to measure the quality of the pupil material. However, the limitations of these instruments are such that one would be unable to interpret small differences in the resulting achievements.

In order to avoid these two difficulties this experiment was planned so that the same teacher should instruct a given class when organized according to two different plans of sectioning. This, necessarily, must be done during successive semesters. This procedure insured the constancy of the teacher, although not necessarily of teaching ability since the ability of a given teacher may vary from semester to semester with different types of class organization. In order that the pupil material might be the same for the two plans of class organization one hundred percent promotion was secured at the middle of the school year. Thus, a teacher who instructed a class as one section during the first semester of this experiment instructed the same pupils during the second semester but with the class divided into two or three sections. Other teachers taught classes organized according to other combinations of sectioning.

This general plan of the experiment makes the semester a variable factor. It is possible that pupils may normally make greater progress during one semester than during the other. Furthermore, the gain of second trial scores over first trial scores is likely to be much greater than the gain of third trial scores over second trial scores simply because the pupils become acquainted with the testing procedure. In order to balance these two variable factors it was necessary to arrange experimental groups in pairs. Thus, corresponding to an experimental group of classes which was taught as a single section during the first semester and as three sections during the second semester, there was another group of classes taught as three sections during the first semester and as a single section during the second semester. In dividing a class into sections the scores yielded by the general intelligence tests were used to secure sections of approximately equivalent pupil material. Six experimental groups of classes were organized as follows:

Group I. Classes taught as a single section during the first semester and as three sections during the second semester.

Group II. Classes taught as three sections during the first semester and as one section during the second semester.

Group III. Classes taught as one section during the first semester and as two sections during the second semester.

Group IV. Classes taught as two sections during the first semester and as one section during the second semester.

Group V. Classes taught as two sections during the first semester and as three sections during the second semester.

Group VI. Classes taught as three sections during the first semester and as two sections during the second semester.

So far as the writer knows, essentially the same methods of instruction and subject-matter were followed in all of these groups. The investigation was confined to Grades II, V, and VII in order to reduce the labor and expense. As these grades are fairly representative of the three divisions of the elementary school, primary, intermediate and grammar, it is not likely that different results would be obtained in the other grades. The number of classes, the total enrollment, and the number of complete records in each experimental group are given in Table I.

TABLE I. NUMBER OF CLASSES, TOTAL ENROLLMENT, AND NUMBER OF COMPLETE RECORDS IN EACH OF THE EXPERIMENTAL GROUPS

Grade		Group I	II	III	IV	V	VI	Total
II	Number of classes	7	4	3	6	7	3	30
	Total enrollment	348	201	138	288	324	162	1461
	Complete records	240	111	103	208	224	89	975
V	Number of classes	2	2	8	4	4	4	24
	Total enrollment	87	92	379	192	196	181	1127
	Complete records	70	72	326	133	157	143	901
VII	Number of classes	3	3	5	5	2		18
	Total enrollment	141	140	244	214	91		830
	Complete records	119	109	186	159	86		659

The data collected. Through the cooperation of Superintendent Peter A. Mortenson of the Chicago Public Schools and of certain principals and teachers, the Bureau of Educational Research carried on this investigation during the school year of 1920-21. Experi-

mental classes were organized in sixteen elementary schools.[1] For measuring the general intelligence of the pupils the Pressey Primer Scale was used in the second grade, and the Illinois General Intelligence Scale in the other two grades. The achievements of the pupils in the second grade were measured by means of the Pressey Scale of Attainment No. 1. In the fifth and seventh grades achievements were measured by Monroe's Standardized Silent Reading Tests, Revised, Monroe's General Survey Scale in Arithmetic, and Buckingham's Problem Scale in Arithmetic, Divisions 1 and 2. The general intelligence tests were given only at the beginning of the experiment, October 11, 1920. Form 1 of the achievement tests was given at this time. Form 2 of the achievement tests was administered at the close of the first semester, February 3, 1921. At the close of the experimental period, May 11, 1921, Form 1 was again given.

The tests were administered by the teachers who also scored the test papers and entered the scores upon individual record cards. This, however, was done only after all of the teachers involved in the experiment had been called together for the purpose of acquainting them with the tests. In this explanation several tests were administered to the teachers in exactly the same way as they were to be administered to the pupils. In addition detailed instructions were supplied to the teachers for all steps of the work. Since no comparisons were made between the scores yielded by tests administered by different teachers it is felt that this procedure in the administration of the tests does not seriously affect the results of the experiment.

Limitations of the experiment to be kept in mind in interpreting the results. A number of conditions must be kept in mind in interpreting the results. In the first place practically all of the teachers who cooperated in the investigation had been accustomed to teaching classes in two sections. A few, perhaps 1 in 20, had taught a class as a single section but, so far as the writer was informed, no teacher had had any experience in instructing a class in three sections. Thus, it is altogether likely that most of the teachers had acquired a technique of instruction which would prove more successful with a class divided into two sections than with a class divided into either one or three sections. Furthermore, there appears to be a prejudice

[1]These sixteen schools were the following: Brown, Dante, Douglas, Fiske, Jenner, Julia Ward Howe, Morse, Otis, Pullman, Scanlan, Shields, Spry, Van Vlissingen, Ward, Wentworth, and West Pullman.

against the division of a class into three sections. Thus, there is introduced a factor which may be expected to produce greater achievements in classes taught as two sections than in classes taught as either one or three sections. The effect of this factor is, however, unknown but it should by all means be recognized in interpreting the results.

The instruments used for measuring the achievements of the pupils do not measure all achievements resulting from instruction. They can be considered to do no more than measure representative samples of the achievements within their respective fields. Outside of silent reading and arithmetic, in which tests were given, there are many important achievements of which no attempt was made to secure direct measurements. It is, of course, possible that the measures of achievements secured correlate closely enough with all other achievements resulting from instruction, that a sufficiently accurate index of all achievements is furnished for judging the relative effectiveness of the instruction in the different experimental groups. However, convincing experimental evidence on the point is wanting and, for this reason, due caution must be exercised in extending the conclusions of this experiment to school subjects other than silent reading and arithmetic, as well as to the more subtle outcomes engendered by the social contacts of the school room.

Finally, it must be remembered that this investigation was carried on in classes enrolling approximately 45 pupils. Hence it does not necessarily follow that the conclusions would apply to classes enrolling 20 to 30 pupils. It is possible that this change in the size of class might produce a complete reversal in the conclusions.

Method of summarizing data. After rejecting records which were incomplete and obviously inaccurate, the scores yielded by an application of a test were combined in a total distribution for each experimental group. Thus, a distribution was formed of the first trial scores made on Monroe's Standardized Silent Reading Tests, Revised, by the group of fifth grade pupils enrolled in "classes taught as a single section during the first semester and as three sections during the second semester." In the same way distributions of scores were formed for each of the experimental groups and for each application of the test. The gain in achievement during the first semester was found by subtracting the average score for the first trial of a test from the average score of the second trial. The gain for the second semester was found by subtracting the average score of the

second trial from that of the third trial. A second measure of gain was secured by following a similar procedure with the median scores but these gains are not given in this report as they were, in general, in agreement with those calculated from the average scores.

In calculating these gains no account was taken of the possible non-equivalence of the different forms of the tests used. In fact no accurate information concerning the equivalence of duplicate forms is available except for Monroe's Standardized Silent Reading Tests, Revised, and for Monroe's General Survey Scale in Arithmetic. The duplicate forms of these two tests have been shown to be approximately equivalent.[2] However, since Form 1 of each test was used twice and the average scores calculated from it were used both as subtrahends and minuends, and since the gain for any plan of sectioning is computed from both semesters the non-equivalence of Forms 1 and 2 of the tests used will not affect the comparisons of gains made in the following tables.

The point scores yielded by the different tests are expressed in terms of different units and from different zero points. Thus before any combination from the results of the different tests can be made it is necessary to express the gains in terms of a common unit. The usual assumption in such cases is that the standard deviation of the distribution of scores represents the same increment of ability for one test as for another. On the basis of this assumption a total distribution for each test was secured by adding the distributions of the six experimental groups within a grade. This was done for the scores secured at each period of testing. The average of the three standard deviations was assumed to represent the same increment of ability for each test and was used as a divisor for reducing the gains to the basis of a common unit. For example, during the first semester the fifth grade pupils in Group I classes made a gain in arithmetic of 23.82 points. During the second semester they made a gain of 21.5 points. The average standard deviation of the arithmetic scores in the fifth grade is 19.65. Using this as a divisor we secure as quotients 1.21 and 1.09. In this manner the entries in Tables II, III and IV were obtained. The two quotients whose calculation was explained are given in Table III.

Tables II, III and IV are similar in structure and are to be read in the same way. The gains for the different experimental groups

[2]Monroe, W. S. Illinois Examination, University of Illinois Bulletin Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. 70 p.

TABLE II. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SECOND GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Test 1			Test 2			Test 3			Test 4			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	240	1.42	.55		.94	.45		1.49	.27		1.37	.71		1.31	.50	
II	111	.90	1.11		.62	.82		.69	1.17		.58	1.18		.70	1.07	
Average		1.16	.83	.33	.78	.64	.15	1.09	.72	.37	.98	.95	.03	1.00	.78	.22
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	103	1.10	.52		.63	.60		1.50	.51		1.39	.47		1.16	.53	
IV	208	.87	1.35		.56	1.03		.48	1.20		.55	1.21		.62	1.20	
Average		.99	.94	.05	.60	.82	-.22	.99	.86	.14	.97	.84	.13	.89	.86	.02
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	224	1.41	.56		.63	.56		1.34	.27		1.53	.62		1.23	.50	
VI	89	.04	1.56		1.01	.42		.28	1.36		.31	1.25		.41	1.15	
Average		.73	1.06	-.34	.82	.49	.33	.81	.82	-.005	.92	.94	-.02	.82	.83	.006

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE III. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE FIFTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	70	.68	−.39		.30	.11		1.21	1.09		.35	.13		.65	.35		.80	.32	
II	72	.55	.09		.58	.37		.28	.75		.24	.40		−.14	.95		.38	.64	
Average		.62	−.15	.77	.44	.24	.20	.75	.92	−.18	.30	.27	.03	.26	.65	−.40	.59	.38	.11
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	326	−.12	.83		.29	.63		.61	.32		.79	.18		.84	.11		.60	.52	
IV	133	.57	−.16		.30	.25		.51	.87		.68	−.05		.25	.24		.58	.29	
Average		.23	.34	−.11	.30	.44	−.15	.56	.60	−.04	.74	.07	.67	.55	.18	.37	.59	.40	.19
		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)		(2)	(3)	
V	157	.23	.69		.45	.66		.54	.29		.45	.20		.79	.16		.62	.50	
VI	143	1.03	−.23		.73	.03		.27	.24		.14	.70		.04	.68		.55	.36	
Average		.63	.23	.40	.59	.35	.25	.42	.27	.14	.30	.45	−.16	.42	.42	−.005	.58	.43	.16

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

TABLE IV. GAINS IN ACHIEVEMENT MADE IN THE DIFFERENT EXPERIMENTAL GROUPS IN THE SEVENTH GRADE*
(GAINS COMPUTED FROM AVERAGE SCORES)

Group	No. of Pupils	Reading Rate			Reading Comprehension			Arithmetic			Arith. Problems 1st Division			Arith. Problems 2nd Division			Average		
		Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.	Gain	Gain	Dif.
		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)		(1)	(3)	
I	119	-.60	.13		.47	.93		.14	.23		.34	-.03		.37	.17		.18	.36	
II	109	.31	-.41		.07	.22		.71	.18		-.20	.58		.07	.78		.24	.34	
Average		-.15	-.14	-.005	.27	.58	-.31	.43	.21	.22	.07	.28	-.21	.22	.48	-.26	.21	.39	-.14
		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)		(1)	(2)	
III	186	-.14	.60		.32	.41		.31	.92		.48	.04		.33	.03		.33	.50	
IV	159	.45	.28		.41	.91		.51	.26		-.07	.53		.23	.08		.38	.52	
Average		.16	.44	-.29	.37	.66	-.30	.41	.59	-.18	.21	.29	-.08	.28	.06	.23	.35	.51	-.15

*The numbers in parentheses in the body of the table indicate the number of sections in which the classes were taught.

are arranged in pairs. In Table II, the gain for Group I on Test 1 when taught in classes of one section is 1.42. When taught in three sections the gain is .55. The gain for Group II classes when taught in one section is .90 and when taught in three sections it is 1.11. The Group I classes were taught in one section during the first semester but the Group II classes were taught in one section during the second semester. This difference in time is largely responsible for the differences in the size of the gains.

Interpretation of results. In interpreting the gains in Tables II, III and IV it is necessary to keep in mind both the constant and variable errors of measurement which are involved in the original data as well as the chance variations in the gains due to sampling. The variable errors of measurement in the original data depend upon the reliability of the tests used. If we assume a coefficient of reliability[3] of .84 for Test 1, it can be shown that the probable variable error of measurement is approximately .25 when expressed in terms of sigma which is the unit used in expressing the gains in Tables II, III, and IV.[4] A probable error of measurement of .25 means that the scores for 50 percent of the pupils involve variable errors which are less than .25. For the other 50 percent the variable errors will be greater than .25. The presence of variable errors of measurement affects the average of the scores as shown by the following formula in which N is the number of scores upon which the average is based.

$$\text{P. E.}_{\text{M average}} = \frac{\text{P. E.}_{\text{M}}}{\sqrt{\text{N}}}$$

Substituting in this formula for Group I, we find the probable error of measurement of the average ($\text{P. E.}_{\text{M average}}$) is .017; for Group II it is .024. The gain 1.42 is the difference between the two averages.

[3]The coefficient of reliability assumed here is probably higher than would be found for this test. When based upon the scores of a single grade, the coefficient of reliability for Monroe's General Survey Scale in Arithmetic is approximately .85. For Monroe's Standardized Silent Reading Test 1, Revised, the coefficients of reliability are approximately .75 for rate and .65 for comprehension. For Test II they are about .08 higher. The reliability of the other tests is not known.

[4]The formula for the probable variable error of measurement is

$$\text{P. E.}_{\text{M}} = .6745\,\sigma\sqrt{1-r_{1I}}$$

In this case $\sigma = 1$.

The probable error of the difference of the two averages is given by the following formula

$$\text{P. E.}_{\text{Dif.}} = \sqrt{\overline{\text{P. E.}_1}^2 + \overline{\text{P. E.}_2}^2}$$

In this formula P. $E._1$ and P. $E._2$ stand for the probable errors of measurement of the two averages whose difference is taken. In this case P. $E._1$ is equal to P. $E._2$ since we have used the average of the standard deviations of the several distributions in reducing the gains to a comparable basis. Applying the above formula, we find that the probable variable error of measurement to be associated with 1.42 is .024 and with .90 is .034. The formula for the probable error of the sum of the two averages is the same as that for their difference. Hence we may calculate the probable error of measurement to be associated with the average gain 1.16 by taking one half of the probable error of measurement of the sum of the two averages. The P. $E._M$ of the average gain 1.16 is .020.

Since the probable variable error of measurement depends only upon the magnitude of the standard deviation of the scores and the number of scores, we will obtain the same result for the gains of these two groups when taught in classes of three sections. The probable variable error of measurement of the difference (.33) may be calculated by the formula given above. It is .028.

This probable variable error of measurement is relatively small in comparison with the gain .33, and in general when an average or difference is three or four times its probable error it can be considered significant. Hence, if we had to consider only the variable errors of measurement we would be justified in asserting that this difference was significant and could not be due to the presence of these errors in our original data. However, it should be remembered that we have been liberal in the estimate of the coefficient of reliability. It is likely that the true value of the probable error is much larger.

Since all gains are expressed in terms of a common unit the probable variable errors of measurement found for the entries under Test 1 will apply also to Tests 2, 3, and 4 provided we assume the same coefficient of reliability for these tests. The probable variable error of measurement of the average is affected by the number of cases from which the average is computed. Hence for the gains made by other groups it will be slightly greater, since the number of scores is smaller

for those groups. In Table III the number of scores in Groups III and IV is slightly larger. Hence a smaller probable variable error of measurement will be found, but for all of the other groups it will be larger than the one which we have considered in detail. In several cases the difference in gains is so small that when compared with the probable variable error of measurement it cannot be considered as significant.

In addition to the variable errors of measurement, it is necessary to consider the chance variations in the gains due to sampling even when the sample has been chosen without bias. The probable error of an average due to sampling is given by the following formula

$$\text{P. E.}_S = .6745 \frac{\sigma_{\text{dist.}}}{\sqrt{N}}$$

Since sigma (σ) has been used as a unit in terms of which the gains are expressed, $\sigma_{\text{dist.}}$ equals 1 for our calculations.[5] In the case of Group I, P. $\text{E.}_S = .044$. The gain 1.42 is the difference between two averages and hence it would be necessary to apply the formula for the probable error of the difference of the two averages. This being done we find that the P. E._S to be applied to the gain (1.42) is .062. In case of Group II, P. $\text{E.}_S = .064$ and for the difference between the two averages it is .090. For the average 1.16, P. $\text{E.}_S = .055$. For the difference .33, P. $\text{E.}_S = .078$.

When we consider the probable error due to sampling (.078) in addition to the probable variable error of measurement (.028) the difference (.33) would probably be significant and indicate a slight superiority in achievement as measured by Test 1 for the pupils taught in classes of one section, provided no other errors could be considered to affect this difference. It is, however, necessary to consider the constant errors of measurement. Their exact magnitude can not be known but their presence is evident. For example, in Table II the gains on Test 1 for Groups I and II when taught as one section are 1.42 and .90 respectively. The gain of 1.42 was made during the first semester and is the difference between the first and second trial scores. The gain of .90 was made during the second semester and is the difference between the second and third trial scores. Due to the pupils becoming acquainted with the tests and

[5]This is not the true value of σ. The variable errors of measurement tend to increase the value of the obtained sigma. The relation is given by the formula

$$\sigma_{\text{true}} = \sigma_{\text{obtained}} \sqrt{r_{11}}$$

the testing procedure, both of these gains involve a constant error. This tends to make the obtained gain larger than the true gain, but as the practice effect of the second trial scores over the first trial scores is larger than that of the third trial over the second trial scores, it is reasonably certain that the gain for Group I (1.42) contains the larger constant error. The gains made by these two groups when taught in classes of three sections are .55 and 1.11. Both of these gains involve a constant error but in this case the larger constant error is found in the gain for Group II. Each of the average gains for these two groups (1.16 and .83) includes a relatively large constant error but the two errors are much more nearly equal than those included in the gains for each group separately. Hence, we are probably justified in considering their difference (.33) to be relatively unaffected by the presence of constant errors in any of our original data.

However, the neutralization of the constant errors which seems plausible, if not probable, in the case we have just considered does not appear to have taken place in a number of the other differences in this group of tables. With the exception of Groups I and II in Table II some of the differences are positive but others are negative for each pair of groups, although it is not impossible that a given plan of sectioning a class might be more effective in one subject than in another. The variations in the signs of the differences do not appear to occur in such a way as to justify this explanation of the negative gains. It is likely that a constant error was introduced in certain groups of scores which was not neutralized in the difference. For example, Group VI is shown by Test 2 to have made a larger gain during the second semester when taught in two sections. Each of the other tests shows a smaller gain for this semester and this we should expect as the gain is the difference between the second and third trial scores. The probable explanation of this condition is that in some way a constant error was introduced in one set of scores yielded by Test 2 for Group VI. An examination of Tables III and IV reveals several similar instances. Hence, we are forced to the conclusion that at least certain sets of scores involve an unknown constant error. The fact that this happened in certain cases tends to make one suspicious of the presence of an unknown constant error in other sets of scores even though evidence of its presence is lacking.

It is perhaps significant that in the case of the differences in gains between classes taught as one section and classes taught in three sections, eight gains are positive while six are negative. The

same situation prevails with respect to the gains made by classes taught in one section when compared with the gains made by classes taught in two sections. For classes taught in two sections compared with classes taught in three sections, we have records only in the second and fifth grades. Four of the differences are positive while five are negative.

Conclusion. The facts presented in Tables II, III, and IV and the errors they include appear to justify the conclusion that there is no evidence of greater achievements being made by pupils when taught in classes organized on the basis of one plan of sectioning than in classes organized on a different plan of sectioning. Since the teachers were more experienced in teaching classes in two sections and probably preferred this plan of organization this condition might appear to mean that the division of classes into two sections was the least efficient of the three plans. However, in the writer's judgment this conclusion is not justified. The most obvious inference, in his opinion, to be drawn from the data of this experiment is that the educational tests used do not yield sufficiently accurate and precise measures of achievement to make possible the determination, under the conditions of this experiment, of the best method of sectioning a class. It is likely that the differences in the gains made during a period of less than a semester are not large. This being the case it is necessary either to extend the experimental period or to secure more precise measures of achievement. The magnitude of the probable variable error of measurement of the difference and also of the probable error due to sampling can be decreased by increasing the number of pupils in the experimental groups, but the constant errors are not affected by any increase in the number of cases. Certain constant errors are neutralized in the differences but, as we have shown, other constant errors which occur in only certain sets of scores were not eliminated. The presence of these constant errors is due to imperfections in the educational tests used. Therefore, it appears that until our instruments for measuring achievements of school children are materially improved we cannot expect such educational experiments as the one described in this report to lead to reliable conclusions.

BULLETIN NO. 12

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE USE OF INTELLIGENCE TESTS AS A BASIS OF SCHOOL ORGANIZATION AND INSTRUCTION

by

CHARLES W. ODELL
Associate, Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1922

TABLE OF CONTENTS

PREFACE

How to use most effectively the information yielded by general intelligence tests is one of the most important questions before the educational world at the present time. Many of our leading educational thinkers are urging that the children in our elementary schools be grouped into grades on the basis of their mental ages and divided into sections within the grade on the basis of intelligence quotients. Other educators maintain that this should not be done. In this monograph Dr. C. W. Odell presents the results of an investigation extending over nearly two years in which he has studied with unusual care certain of the questions involved in the proposal that we reorganize our schools on the basis of the results yielded by general intelligence tests. Because the questions studied are highly important it is felt that a somewhat detailed report is justified. In order to assist the reader in understanding the experiment the organization of the experimental schools has been described in detail.

This investigation was undertaken at the invitation of Superintendent Peter A. Mortenson of Chicago. Its execution was made possible by the cooperation of Assistant Superintendent A. B. Wight and of certain principals and teachers in the Chicago public schools. To all who have cooperated in the course of the investigation the Bureau of Educational Research desires to acknowledge its indebtedness.

WALTER S. MONROE, *Director.*

November 10, 1922.

THE USE OF INTELLIGENCE TESTS AS A BASIS OF SCHOOL ORGANIZATION AND INSTRUCTION

CHAPTER I

THE PLAN AND CONDUCT OF THE EXPERIMENT

The Problem. The experiment described in this bulletin was carried on in eight elementary schools in the city of Chicago. It was an attempt to answer the following question: What is the effect upon the efficiency of elementary schools of promoting and classifying pupils chiefly upon their mental ages and intelligence quotients as determined by group intelligence tests rather than according to the traditional method? It is recognized that this is really a double problem involving the question of a flexible system of promotion and classification upon any basis as compared with a non-flexible system, and also the question of using the results of group intelligence tests rather than some other basis for promotion and classification. The justification for combining these two questions is that the use of the results from group intelligence tests for the purposes mentioned above necessarily involves a flexible system and hence the two questions may be considered as one from the standpoint of practical school administration.

Definition of terms used in statement of problem. The "efficiency" of a school is the ratio of its output to the investment, or $\frac{\text{output}}{\text{investment}}$. The output or return upon the investment is measured in terms of the achievements of the pupils and their rates of progress through the school system. The investment, as the term is used above, includes not only what might strictly be called investment but also the factors which affect the manner in which the investment proper is used. The "promoting" of pupils refers to their advancement from one half-grade to another. The "classifying" of pupils refers to their placement in the fast, average and slow sections into which each half-grade in the experimental schools was divided. The word "chiefly" is used in the statement of the problem because the information derived from group intelligence tests was supple-

mented by other data. The "traditional method" refers to the method of placement used in the group of control schools. According to this method, promotion is determined by the pupil's final mark, which is usually a composite of the mark that he receives upon the final examination and that given by the teacher for his work during the term. In some cases the promotion indicated by the pupil's final mark is modified by the principal's opinion of his work or ability or by such considerations as chronological age, length of time already spent in the grade, number of pupils in the room, etc. This is the method which has been and still is the prevailing practice in the elementary schools of this country.

Scope of study. This study was confined to elementary schools having sixteen[1] or more teachers, which were organized in sixteen half-grades and in which pupils were promoted semi-annually. These schools were divided into an experimental and a control group of four each by Assistant Superintendent A. B. Wight. In making this selection, Mr. Wight endeavored to choose two groups of schools[2] in which the investment factors should be approximately equal at the beginning of the experiment.[3] Except in the plan of organization of the experimental schools, nothing was done to cause any change.

This investigation was rather strictly limited to the study of the effect upon the "efficiency" of certain elementary schools of promoting and classifying pupils chiefly according to the data derived from the use of group tests of intelligence. There was no consideration of the desirability of discovering and segregating for instructional purposes pupils of varying degrees of ability, except from the standpoint of their rates of progress and achievements in school. Neither was there any assumption that the plan used, which provided that pupils of different degrees of ability should complete the same course of study at different rates of progress,

---

[1]In one of the control schools there were only fourteen teachers in charge of pupils who actually participated in the project.

[2]The experimental schools were the Armour, Franklin, Holden and Moseley. In the control group were the Alcott, Greene, Mark Sheridan and Webster.

[3]A more complete comparison of the investment factors in the two groups of schools may be found in the dissertation of the same title and by the same writer as this bulletin. This dissertation is on file in the library of the University of Illinois.

was superior to a plan providing that the different groups of pupils should cover different courses of study.

In this experiment the measurement of the achievements of the pupils was, with certain minor exceptions, limited to arithmetic and reading. As these are two of the most fundamental subjects studied in the elementary school, they were considered a fairly good measure of total achievement.

The general plan of the experiment. The experiment began in the autumn of 1920 and continued until the summer of 1922. As Table I shows, both intelligence and subject-matter tests were administered to the pupils of both groups of schools once each semester except that at the last testing only subject-matter tests were used. Also certain other data that seemed pertinent to the investigation were collected at each time of testing. After the first testing period the pupils of the experimental schools were promoted and classified upon the basis of the test results and the other data which had been obtained. After each of the later testing periods such adjustments were made as seemed advisable. No direct use was made of the test results or other data in the control schools. There was some opportunity for the teachers of this group of schools to make a more or less indirect use of the test results but they did not do so to a degree sufficient to affect the results of the experiment. By thus measuring the abilities and achievements of the pupils of the two groups of schools near the beginning and end of each semester the resulting data afforded a basis for comparing the effect of promoting and classifying pupils chiefly upon the results obtained from the use of group tests of intelligence with that of promoting pupils according to the traditional method.

The first tests were given in November, 1920, and the results used in promoting and classifying the pupils for the second semester of 1920-21. The next testing occurred in May, 1921, and furnished the basis for the placement of the pupils for the following September. All new entrants were tested in September and assigned to their grades and sections as soon thereafter as possible. A general testing occurred again in December and was followed by the placement of the pupils for the second semester of 1921-22. The final testing was in May, 1922. In addition to these general testing

periods, small groups of absentees and new entrants were tested from time to time as seemed best.

The tests were in all cases given by the regular teachers who had been prepared for this work by a careful program of meetings with discussion. The teachers were also given very detailed directions. Rather extensive visiting by the writer while the tests were being administered showed that this program of preparation secured fairly uniform and correct procedure. Part of the scoring of the tests was done by the teachers and part by clerks in Assistant Superintendent Wight's office. The scoring was also checked sufficiently by the writer to warrant the belief that it was fairly accurate. Most of the errors which were found were so small that they had no effect upon the placement of pupils. In the tabulation of test and other data the positive and negative errors balanced each other so as to leave no sensible inaccuracy in the medians and other measures computed.

The data collected. Table I shows the intelligence and achievement tests used at each date of testing. The scores made upon these tests were translated into mental[4] or achievement[5] ages, as the case might be, and then further into intelligence[6] and achievement[7] quotients. The mental ages and intelligence quotients used

---

[4]Mental age is a term used to express the amount of intelligence possessed by an individual. The average score made upon an intelligence test by a large number of unselected children of any one given chronological age is said to be equal to a mental age of the given number of years. Thus, if on a given test the average score of six-year-olds is 25 points and that of seven-year-olds is 30 points, a score of 25 points may be transmuted into a mental age of six years and one of 30 points into one of seven years. It is abbreviated M.A.

[5]Achievement age is used to express the amount achieved by an individual on a subject-matter test. The average score made by the children of a single mental age is taken to equal an achievement age of the same number of years. Thus, if the average score of children of the mental age of ten years is 56 points, that score may be changed into an achievement age—abbreviated A.A.—of ten years.

[6]The intelligence quotient is the ratio of the mental age to the chronological age, or M.A. divided by C.A. It is conventionally carried to two places and written without the decimal point. Thus a child who has a mental age of ten years and is eight years old has an intelligence quotient of 10 divided by 8, or 125. It is abbreviated I.Q.

[7]The achievement quotient—abbreviated A.Q.—is the ratio of the achievement age to the mental age, or A.A. divided by M.A. It is written similarly to the I.Q. Thus a child whose achievement age is nine years and whose mental age is ten years has an achievement quotient of 9 divided by 10, or 90.

TABLE I. THE INTELLIGENCE AND ACHIEVEMENT TESTS USED IN THIS EXPERIMENT

Date of Testing	Grades	Intelligence Tests	Grades	Achievement Tests
Nov. 1920	IB-IIIB	Indiana University Primer Scale (Pressey Primer)	IA	Indiana University First Grade Reading Vocabulary Test, Form A
	IB-IIIB	Dearborn Group Tests of Intelligence	IIB-IIIB	Indiana University Scale of Attainment No. 1, Form A
	IIIA-VIIIA	National Intelligence Tests, Scale A, Form I	 IIIA-VIB VIA-VIIIA	Monroe's Standardized Silent Reading Tests Form I, Test 1 Form I, Test 2
	IIIA-VIIIA	Illinois General Intelligence Scale, Form I	 IIIA-VIB VIA-VIIIA	Monroe's General Survey Scale in Arithmetic Form I, Scale 1 Form I, Scale 2
May 1921	IB-IIIB	Indiana University Primer Scale		The same tests were used as in November, 1920, except that Form B of the two Indiana Tests and Form II of the Monroe Tests were used.
	IIIA-VIIIA	Illinois General Intelligence Scale, Form II		
Sept. 1921	IB	Kingsbury Primary Group Intelligence Scale, Form A		The same tests were used as in May, 1921
	IA-IIIB	Indiana University Primer Scale		
	IIIA-VIIIA	Illinois General Intelligence Scale, Form II		
Dec. 1921	IB-IIIB	Myers Mental Measure		The same tests were used as in November, 1920, except that Form III of the Monroe Tests was used.
	IIIA-VIIIA	Illinois General Intelligence Scale, Form I		
May 1922		None used.		The same tests were used as in November, 1920, except that Form B of the Indiana Vocabulary Test and Revised Form A of the Scale of Attainment No. 1 were used.
			VIB-VIIIA	Omnibus Test*

*See Appendix C.

at the first time of testing were based upon the average of the two mental tests given at that time.

The other items of information called for by the individual record cards used in this project were as follows: name, building, date of birth, sex, date of testing, school grade, chronological age, teacher's estimate, average school mark, attendance, and health mark.

The teacher's estimate was an opinion as to the general capacity of the pupil regardless of whether this capacity was actually displayed in regular school work or not. This estimate was expressed

in terms of the following five marks and the teachers were instructed to make their distributions accord fairly closely with that given below:

S or 5—superior	—	5 to 10	percent
E or 4—excellent	—	20	percent
G or 3—good	—	40 to 50	percent
F or 2—fair	—	20	percent
P or 1—poor	—	5 to 10	percent

The average school mark was the average of the pupil's marks upon the seven most important subjects of the course of study. It was based upon the grades on the monthly reports issued during the current semester previous to the date of testing and was expressed in terms of the same five marks that were used for the teachers' estimates.

Attendance was given as the percent of school days from the beginning of the semester to the date of testing during which the pupil was present. The health mark was the teacher's opinion of the general health of the pupil and was expressed in terms of the same five marks that were used for teachers' estimates and average school marks.

At the time of the first testing, the published norms and data for the transmutation of point scores upon the tests used into mental and achievement ages were in most cases based on a number of pupils not much larger, or even actually smaller, than the number taking the tests in this experiment. Hence it was decided that in the case of most of the tests used, the norms and tables for transmutation should be based upon the data obtained in this project.[8] The exceptions to this decision were the Illinois Examination, including the Illinois General Intelligence Scale and Monroe's Arithmetic and Reading Tests, and the Myers Mental Measure. These exceptions were made because in the case of the Illinois Examination scores from about fifty thousand pupils were available,[9] and in that of the Myers Mental Measure scores from about fifteen thousand pupils.[10]

---

[8]See complete dissertation for these transmutation tables and their derivation.

[9]Monroe, W. S. A Report of the Use of the Illinois Examination, Form 1, with 49,500 Pupils. Insert of School and Home Education, March, 1921. 8p.

[10]Myers, C. E. and G. C. Measuring Minds. New York; Newson, 1921, p. 23-4.

Principles of promotion and classification used at the first placement of the pupils. It was necessary to lay down certain principles which should be followed in the placement of the pupils, with the understanding that there would be need for exceptions in the cases of certain individuals. The inadvisability of following set rules too closely was due to several facts. Such procedure would result in entirely too great a change in the placement of some pupils. Moreover, the data secured from the tests and from other sources could not be relied upon as being absolutely accurate, and in some cases were so conflicting that disagreements between any detailed principles laid down were sure to occur. In view of these facts it must be understood that the principles enumerated below were not adhered to absolutely and that there were exceptions of many sorts that it is impracticable to list. The principles formulated for the first placement of pupils, which was for February, 1921, are given below.

I. The use of the data derived from the intelligence tests.

1. The chief bases of placement were the mental ages and intelligence quotients. The mental ages were used to determine the half-grades in which the pupils should be placed, and the intelligence quotients to determine the sections, subject to such modifications as may be given in II.
2. In general, the mental age norm for each half-grade was the median mental age of this half-grade group for the experimental schools. If the median mental age of a particular half-grade group in any one school was distinctly above or below the median of the four schools, a rough average of the two medians was used. This was done because it was considered desirable to make some progress toward reducing the range of ability within a given half-grade group for the experimental schools, but not to do so without regard to the ability actually found in the half-grades of the several schools as they were at the beginning of the experiment. Rather wide mental age limits were used for each half-grade group, with the expectation that as the experiment progressed they would be narrowed.

3. In general, demotion was recommended only when a pupil's mental age was at least two years below the median of the grade in which he was found, and extra promotion only when it was at least several months higher than the median of the grade in which extra promotion would place him. In no case was a pupil recommended for skipping more than two semesters' work nor for being demoted more than one.
4. Pupils whose intelligence quotients were above 110 were usually placed in the fast sections, those with I. Q.'s below 85 in the slow sections and the remainder[11] regularly composed the average sections. However, in many cases pupils' mental ages were several months above the medians of the half-grades in which normal promotion would place them, while their I. Q.'s were below 85. In such cases they were usually given normal promotion to the average section of the next half-grade. Similar exceptions were made in connection with other ranges of mental ages and intelligence quotients.
5. In some cases where the mental ages and intelligence quotients were rather low, it appeared probable that the pupils had either misunderstood directions upon one of the two intelligence tests or, had not, through some other cause, done themselves justice upon one of them. In such cases their scores on the other test were given more than half weight in determining their placement.

II. The use of the other data obtained.

1. Most of the other items recorded upon the individual record cards were given consideration. Low teachers' estimates and average school marks rarely prevented promotion in cases where the mental ages seemed to warrant it. In doubtful cases the question of whether

[11]Inasmuch as the intelligence quotients obtained in this project were derived from group intelligence tests they had a somewhat greater spread than those derived from individual tests. Therefore the percent of pupils with I.Q.'s from 85 to 110 was somewhat less than that usually found between 90 and 110 when individual tests are used.

a single or a double promotion should be given was, however, frequently decided by the teachers' estimates and average school marks.

2. On the other hand, even though their mental ages were low enough to merit demotion, very few pupils were failed whose teachers' estimates and average school marks were "good" or better, and not very many were failed if either one of the two was this high.
3. In making use of the teachers' estimates of capacity and the average school marks it was found that those of some teachers ran much higher than those of others in cases where the mental and achievement ages of the two groups of pupils showed little difference. That is, a teacher's estimate or school mark of "fair," for instance, given by one teacher might be fully equal to one of "good" given by another. In making use of these two items an allowance was made for this fact.
4. In cases of marked disagreement between the evidence afforded by the test data and that given by the teachers' estimates and school marks, the achievement test scores were frequently the deciding factor in placement.
5. In doubtful cases the teachers' estimates of health sometimes determined placement, but were not a major factor.

III. Pupils not classified.

1. Since all promoted VIIIB and VIIIA pupils would leave the schools concerned before the close of the experiment, no recommendations were made in the case of any pupils of these grades.
2. Pupils in open-air and ungraded rooms were recommended for promotion and classification according to the same principles used for the other pupils, but it was not expected that they would actually be placed in exact agreement with the recommendations.

An illustration of the application of these principles of promotion and classification. In order to illustrate the actual application of these principles, the following sample taken from the

TABLE II. SAMPLE OF THE PROMOTION LISTS MADE OUT FOR THE BEGINNING OF THE SECOND SEMESTER OF 1920–21

Pupil Number	M.A.	I.Q.	A.A.	T.E.*	School Mark	Health Mark	Placement of Pupil
1	6.6	73	8–8	1	1.8	3	IIIA slow
2	7.7	66	7–10	2	1.7	3	IVB slow
3	11.3	110	10–8	4	3.8	3	VB average
4	8.0	85	9–6	2	2.7	4	IVB fast
5	11.3	100	11–10	3	3.0	4	VB average
6	9.1	106	9–8	3	3.5	3	IVA average
7	9.2	95	8–8	3	3.1	3	IVA average
8	10.2	105	9–8	3	3.0	3	IVA average
9	10.6	78	11–10	3	3.0	4	IVA slow
10	8.8	71	10–0	2	2.0	1	IV B slow
11	4.7	40	7–4	1	1.0	3	IIIA slow
12	10.0	72	8–10	2	2.0	3	IVA slow
13	12.3	109	11–6	3	3.0	3	VB fast
14	10.4	91	11–2	3	3.0	4	IVA average
15	10.4	80	8–0	3	2.8	2	IVA slow
16	8.7	106	9–0	2	2.7	4	IVA average
17	10.2	69	13–0	3	3.0	4	VB average
18	9.2	106	8–6	3	3.2	3	IVA average
19	8.6	85	8–4	2	1.5	1	IVB slow
20	11.2	100	11–10	4	4.0	4	VB average

*Teacher's Estimate.

lists actually made out is given and discussed. These lists were later submitted to the principals and teachers, as has been mentioned previously, and any changes that seemed best were made.

Since the sample in Table II is a portion of the list for the pupils who were in the IVB grade during the first semester of 1920–21, the median mental and achievement ages[12] for the half-grade groups into which IVB pupils were likely to be placed are given below.

Grade	IIIA	IVB	IVA	VB
Mental Age	8.1	9.6	9.8	10.8
Achievement Age	7–4	9–1	9–2	10–6

The mental ages of Nos. 1 and 11 were so low that it was evident they should be demoted. This was corroborated by the low teachers' estimates and average school marks given them. Their I. Q.'s clearly indicated that they belonged in the slow section. Nos. 2, 10 and 19 had mental ages considerably below the IVB median and I. Q.'s of 85 or below. As their teachers' estimates and school marks were also fairly low they were kept in their grade

[12]See Table III.

and placed in the slow section. No. 12 had a mental age above the IVB median, but an I. Q. of only 72, so he was recommended for the IVA slow section. Although No. 4's mental age and I. Q. were low enough to indicate that he belonged in the IVB slow section his fairly good achievement age and his school mark of 2.7 resulted in his being placed in the fast section of that grade. This was done to prevent him from having to repeat work during the whole of the next semester and with the expectation that he would soon drop back into an average or slow section. The mental ages of Nos. 9 and 15 seemed to entitle them to extra promotion but as their teachers' estimates and school marks were only about average they received merely normal promotion into the IVA grade. Because of their low I. Q.'s they were placed in the slow section. In the case of the six pupils placed in the average section of IVA there was little doubt as to where they belonged except that No. 16 had a mental age almost a year below the IVA median. His rather high I. Q. and average school mark led to the decision not to prevent his advancement. Nos. 8 and 14 might have been considered for extra promotion had their teachers' estimates and school marks been higher. Nos. 3 and 20 were clearly entitled to extra promotion on the basis of all the data and No. 5 was only slightly less deserving. The I.Q. of the first would have caused his placement in the fast section but it happened there were not enough pupils in the school of similar ability to justify the formation of a fast section in grade VB. Therefore all three were placed in the average section of that grade. No. 17 was also given extra promotion. In his case a chronological age of almost 16 years and a high score on the achievement tests were potent clauses. For the same reason he was placed in the average rather than the slow section, although his I. Q. was only 69. No. 13 would probably have been given two semesters of extra promotion instead of one except for the fact that his teacher's estimate and school mark were only 3. As it was he was given one extra promotion and placed in the fast section of the grade.

Supplementary principles of promotion and classification used at the second and third periods of placement of the pupils. At the second and third periods of placement—that is, for September, 1921, and February, 1922,—a majority of the pupils

placed at the beginning of the second semester of 1920-21 received normal promotion into the next half-grade and remained in the corresponding section. Unless the new data clearly indicated that the pupil had been placed improperly in February, 1921, this course was followed. Certain additional principles were adopted to care for those pupils who seemed to have been improperly placed. These principles were as follows:

1. In the cases of a number of the pupils given extra promotion at the beginning of the previous semester, their school marks and achievement ages following this promotion did not appear to justify it. If, however, their mental ages as shown by the later testing were high enough to justify their retaining the extra promotion given and also receiving normal promotion at the later date, such promotion was usually given. This was done on the assumption that after skipping the work of one or more semesters it might require more than one semester for them to "find themselves."
2. Pupils previously promoted or placed in fast sections despite their low school marks were failed if their school marks still continued to be unsatisfactory.[13]
3. Many pupils who had received only a part of the extra promotion that they seemed to deserve in February, 1921,[14] were given further extra promotion, if their later scores justified so doing.
4. In cases where the test scores of pupils varied greatly from those made at the previous testing period or periods, and the other evidence did not agree with one score more than the other, the scores were roughly averaged to provide the basis for placement.

The final placement of the pupils. The writer made out his recommendations for placement, basing them upon the principles listed above, some two or three weeks before the end of the semester. The lists were then submitted to the principals and teachers concerned for their consideration, and finally put into

[13]Most of these seemed to be cases of laziness and lack of study.

[14]These pupils had received only a part of their extra promotion in order to lessen the amount of work skipped at one time and thus make their advance easier.

effect. In two of the four schools the recommendations were discussed individually, but in the other two this was not done, as the principals of those schools wished to make as complete a change as possible from the traditional method of procedure. The changes made as a result of this consideration amounted to about one percent of the total number of recommendations made. These changes were often due to the fact that a longer acquaintance with certain pupils caused the teachers to wish to revise the estimates of capacity or school marks which had been reported some time previously. Sometimes, however, the changes made represented a yielding on the part of the writer of his judgment, based largely upon the test results, to that of the principal or teacher, which was based upon the actual school work of the pupils and upon personal contact with them. In a few cases a change was made in order to place the pupil under a certain teacher so that he would be separated from a group of classmates.

In planning this whole experiment and in formulating and using the principles of promotion and classification those in charge of the experiment were guided by the desire to do a practicable piece of work. That is to say, they wished to use a procedure which the average school administrator or supervisor would be willing and able to make use of in his own school. It was partly because of this desire that more thoroughgoing changes were not made in the placement of the pupils, especially after the first period of testing. It is true that some public school superintendents have carried out considerably more radical plans of reclassification than the one used in this experiment but it was believed that a plan that might be followed by a more conservative educator would be more worth while.

CHAPTER II

CONDITIONS AT THE BEGINNING OF THE EXPERIMENT

Chronological age-grade placement in the two groups of schools. The chronological age-grade situation in November, 1920, showed that the retardation in the experimental schools was somewhat greater than that in the control schools. The median age of the pupils of the experimental schools averaged, grade for grade, two-tenths of a year more than that for the other group. In only two of the half-grades was it lower. The percents of pupils accelerated, normally placed, and retarded were 9, 18 and 73, respectively, in the experimental schools as compared with 10, 21 and 69 in the control schools. These figures are based upon the Chicago standard of normal progress, which is that a pupil should be from six to six and one-half years of age in grade IB, six and one-half to seven in grade IA and so on up. The average amount of retardation per pupil[1] was 1.14 years for the experimental schools and .96 year for the control schools. Assuming that pupils had entered the two groups of schools at the same average age, which the writer believes was the case, it is evident that the pupils in the control schools at the beginning of the experiment had made somewhat more rapid progress than had those in the experimental schools.

Mental age and school placement in the experimental and control schools. As may be seen from Table III, the median mental ages in all except three of the half-grades were higher in the control than in the experimental schools. The average difference was slightly over one-half year of mental age. This difference was found in spite of the fact just mentioned above that the pupils of the control schools were grade by grade about two-tenths of a year younger than those of the other group.

[1]The average amount of retardation was computed as follows: The number of pupils accelerated one-half year was multiplied by one-half, the number accelerated one year by one, and so on. The same process was carried out for those retarded and the sum found for each group of pupils. As the total number of years of retardation was greater than the total of acceleration, the latter was subtracted to give the net total of retardation. This was divided by the total number of pupils.

TABLE III. GRADE MEDIAN MENTAL AND ACHIEVEMENT AGES, INTELLIGENCE AND ACHIEVEMENT QUOTIENTS OF THE EXPERIMENTAL AND CONTROL SCHOOLS, NOVEMBER, 1920

Grade	Mental Ages		Intelligence Quotients		Achievement Ages		Achievement Quotients	
	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.	Exp.	Cont.
IB	6.0	6.4	88	98				
IA	7.5	7.7	100	105	7–3	7–8	97	103
IIB	7.7	8.7	95	110	6–9	7–8	89	87
IIA	8.2	9.1	92	107	8–5	9–0	104	102
IIIB	9.0	8.6	100	96	10–1	9–10	113	113
IIIA	8.1*	8.3*	77*	84*	7–4*	7–10*	110*	110*
IVB	9.6	10.0	92	102	9–1	9–0	110	102
IVA	9.8	10.4	92	100	9–2	9–6	104	105
VB	10.8	10.7	93	98	10–6	10–7	106	108
VA	10.0	11.2	85	100	10–1	11–4	112	106
VIB	11.8	11.8	97	96	10–9	11–8	100	105
VIA	12.6	13.2	101	103	11–3	12–1	105	104
VIIB	12.5	14.4	99	112	11–6	13–0	105	100
VIIA	13.2	14.0	102	106	11–11	13–2	107	109
VIIIB	14.2	14.8	104	110	12–2	14–4	99	109
VIIIA	15.4	15.8	111	114	13–7	14–7	99	105
All	9.1	9.7	94	103	10–0	10–8	104	103

*The low M. A.'s and I. Q.'s found in grade IIIA were doubtless due to the fact that the Illinois General Intelligence Scale requires a degree of reading ability somewhat above that possessed by most IIIA pupils.

**In grades IIIA to VIIIA a composite achievement age was obtained by averaging the achievement ages upon Monroe's arithmetic scale and in comprehension and rate upon his reading test. In obtaining this average each of the three was given equal weight. The same procedure was followed in the case of the achievement quotients.

The inter-quartile ranges of the various grades were also computed. These showed an average range of two and one-tenth years for the experimental schools and two and two-tenths years for the control schools. Thus it appears that the grade groups in the experimental schools were slightly more homogeneous than those in the control schools. Similarly, a slight advantage was shown by the coefficients of correlation of mental age and grade placement. These were $.84 \pm .01$[2] for the experimental schools and $.82 \pm .01$ for the control schools.

The intelligence quotients of the two groups of schools. Probably the best basis of comparing the mentality of the pupils of the two groups of schools is that of their intelligence quotients. Table III shows the medians for the two groups of schools. In all grades except IIIB and VIB the median I. Q.'s of the experimental

[2] For convenience all probable errors smaller than .01 are given as .01

schools were lower than those of the control schools. The average difference was nine points, the medians for all grades combined being 94 and 103. This of course agrees with the fact just noted that the mental ages of the pupils of the experimental schools were lower, although their chronological ages were higher, than those of the other group.

The extent to which the data derived from the tests afforded a true comparison of the quality of the pupil material of the two groups of schools depends upon the reliability of the tests used and the similarity of testing conditions in the two groups of schools. As is shown in Appendix B, the reliability of the intelligence tests was only fairly high, but there is no reason to think that the degree of reliability was different in the two groups of schools. Moreover, as has been stated in Chapter I, the writer's rather extensive visiting while the tests were being given and his examination of the test booklets after they had been scored afforded fairly reliable grounds for believing that there were no essential differences in the administration of the tests in the experimental and in the control schools.

The use of the control schools as a check group upon the experimental schools. If we assume that the difference in the amount and degree of intelligence found by the use of the intelligence tests was reliable, the question remains as to whether this difference was so great that the control schools could not be used as a valid check upon the experimental schools. A definite answer to this question cannot be given. Such data as are available concerning the mentality of pupils of different school systems appear to show that an average difference of about seven months of mental age or nine points I. Q. is not unusual. Probably the most extensive data available upon this point are those obtained from the use of the Illinois General Intelligence Scale.[3] This scale was given to the pupils of ten cities and nine counties in the autumn of 1920. It was found that the differences between the median mental ages of the various grades of the single cities and counties concerned and the general medians for the corresponding grades were four months or more in 50 percent of the cases. The largest difference was one year and three months. In terms of the I. Q. 50

[3]Monroe, W. S. A Report of the Use of the Illinois Examination, Form 1, with 49,500 Pupils. Insert of School and Home Education, March, 1921. 8p.

percent of the differences exceeded four points, the greatest being nineteen points. Differences as large as the average difference between the two groups of schools in this experiment were found in about one-sixth of the cases. Moreover, it must be remembered that differences between the individual members of a group and the group median are, on the average, much less than the differences between the individual members of the group. On the other hand, the differences in this experiment were based upon the average scores from two tests and therefore would probably tend to be smaller than those based upon a single score. This latter factor would not more than balance the one mentioned in the previous sentence, however, and probably would not even do that. Therefore the writer feels justified in the opinion that the difference in mentality found to exist between the two groups of schools was not so great but that the question referred to above can be answered affirmatively, provided that this difference was measured and taken account of in interpreting the results of the experiment.

The achievements of the two groups of schools. Table III also contains the median achievement ages for the various grades. It shows that the control schools were superior in all of the half-grades except IIIB and IVB. This average superiority was about eight months of achievement age, which is enough to indicate a decided superiority in pupil achievement on the part of the control schools.

The achievement quotients, however, are really more significant measures than are the achievement ages. It is evident from Table III that on the whole the relation of achievement to capacity, in so far as the tests used measured this relation, was practically the same in the two groups of schools. This would be inferred from a study of the mental and achievement ages. Such a comparison shows that the superiority of the control schools in achievement was just about the same as their superiority in intelligence. Thus from this standpoint the two groups of schools were capitalizing the capacities of their pupils almost equally in so far as the achievements measured were concerned.

The correlation of achievement with intelligence in the two groups of schools. It is a belief of many educators that the achievements of pupils should be as closely related to their capaci-

ties as possible and that the degree to which this relation holds is a measure of the success of the school in adapting its work to the individual pupils. The achievement quotient measures this from one standpoint, but it may also be measured by computing the correlation of achievement with intelligence. For all grades combined the coefficients of correlation between absolute achievement and intelligence scores was .68±.01 for the experimental schools and .60±.01 for the control schools. That is to say, the experimental schools were securing achievement more nearly in proportion to pupil capacity than were the control schools.

Another measure of the relation of achievement to intelligence may be obtained by computing the median achievement quotients for pupils of different levels of intelligence. This measure is based upon the assumption that the school should secure from all pupils the best work of which they are capable, and if it can not do this it should approach the standard as nearly for pupils of one level of intelligence as for those of another.

Table IV presents the median achievement quotients for the pupils of different levels of intelligence in the two groups of schools. A study of this table reveals the fact that in both groups of schools the inferior pupils were achieving more in relation to their capacity than were the superior pupils, but that this tendency was somewhat

TABLE IV. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE, NOVEMBER, 1920

Intelligence Quotient*	Achievement Quotients	
	Experimental	Control
150–59	104	98
140–	95	95
130–	102	98
120–	102	102
110–	101	101
100–	102	102
90–	104	104
80–	103	105
70–	107	110
60–	113	114
50–	109	125
All	105	104

*Only those levels of intelligence were included that had a sufficient number of cases to give fairly reliable medians.

less marked in the experimental schools. Further evidence to the same effect may be obtained from a comparison of the coefficients of correlation of the achievement and intelligence quotients. These were $-.16\pm.01$ for the experimental schools and $-.28\pm.01$ for the control schools.

Teachers' estimates of capacity, average school marks and estimates of health, in the two groups of schools. The pupil material of the two groups of schools may also be compared by means of the teachers' estimates, average school marks and health estimates. It is true that these measures are relatively subjective, but as there were almost one hundred teachers in each group of schools and as there was no apparent selection which would make one group of teachers more able to judge pupils than the other, these measures were probably fairly comparable for the two groups of schools. Taking the medians for all pupils, the teachers' estimates for the control group were two-tenths higher, the average school marks three-tenths higher, and the estimates of health two-tenths. Considering the three items together, the half-grade medians of the control schools were higher in about 50 percent of the cases, those of the experimental schools in only about 25 percent, and the two were equal in about 25 percent. The evidence afforded by these items is of value chiefly because it corroborates that obtained from the intelligence and achievement test results.

Summary. The differences found to exist between the experimental and the control schools in November, 1920, at the beginning of the experiment, were on the whole large enough not to be neglected as due to chance or as of no consequence, but were not large enough to invalidate the use of the two groups of schools in this experiment. When contrasted with the control schools the experimental schools exhibited the following differences:

1. .18 year greater retardation based on chronological age (1.14 years — .96 year)
2. .6 year lower median mental age (9.7 years — 9.1 years)
3. 9 points lower median I. Q. (103 — 94)

4. 8 months lower median achievement age in reading and arithmetic (10 years 8 months — 10 years)
5. 1 point higher achievement quotient in reading and arithmetic (104 — 103)
6. .08 higher correlation of achievement with intelligence. (.68 — .60)
7. .2 lower median teachers' estimate, school mark and estimate of health, averaged (3.3 — 3.1)

CHAPTER III

THE EFFICIENCY OF THE TWO GROUPS OF SCHOOLS AS MEASURED BY THE RATES OF PROGRESS OF THE PUPILS

In Chapter I "efficiency" was defined as the ratio of the output to the investment. The output to be measured was limited to the achievements of the pupils and their rates of progress. As was stated, the various factors constituting investment were all approximately constant except that of the mental abilities of the pupil material. Therefore, the "efficiency" of the experimental and the control schools might be measured in terms of the ratios of the achievements of the pupils and their rates of progress to their mental abilities. This chapter presents the data dealing with the rates of progress of the pupils, and the relation of these rates to their mental abilities.

The promotion and classification of the pupils for February, 1921. The information obtained from the testing in the four experimental schools in November, 1920, formed the chief basis for the placement of pupils for the succeeding semester, the second of 1920-21. This placement was made by the writer, following the principles of promotion and classification enumerated in Chapter I. The first half of Table V shows the percents of pupils in each grade of the experimental schools gaining or losing various amounts as a result of this placement. It is to be interpreted as follows, using grade IIB as an example: 7 percent of the pupils in grade IIB during the first semester of 1920-21 were demoted one semester, that is, were placed in IA; 26 percent were failed and remained in IIB; 64 percent received regular promotion of one semester into IIA and 2 percent received an extra promotion and thus entered IIIB. It will be seen that on the whole the placement of the pupils in the experimental schools involved many more demotions and failures than extra promotions and that the percent of the pupils given normal promotion was not as great as is usual in school systems. Only 58 percent of the pupils

TABLE V. PERCENTS OF THE PUPILS PROMOTED, FAILED OR DEMOTED THE GIVEN NUMBER OF SEMESTERS AT THE CLOSE OF THE FIRST SEMESTER OF 1920-21

First Semester Grade	Experimental Schools Semesters Gained or Lost					Control Schools Semesters Gained or Lost					
	−1*	0	+1	+2	+3	−1	0	+1	+2	+3	+4
IB		48	49	3	1		31	68	1		.2
IA	17	17	48	9	9		34	60	7		
IIB	7	26	64	2		2	9	80	9		
IIA	1	19	70	10		.4	16	83	.4		
IIIB	2	12	75	9	1		14	82	4	1	
IIIA	14	31	50	3	2		32	67	1		
IVB	4	19	57	13	8		10	82	7		
IVA	6	19	63	6	5	1	7	88	5		
VB	4	16	53	24	3		16	79	5		
VA	16	22	55	4	1		4	95	1		
VIB	10	28	43	15	4		11	87	2		
VIA	11	18	59	12			16	79	5		
VIIB	17	23	44	16			4	96	1		
VIIA	8	13	75	5			18	78	4		
All	6	26	58	8	2	.3	17	79	4	.03	.03

*−1 denotes one semester lost through demotion, 0 failure, +1 normal promotion, +2 one extra promotion, etc.

received normal promotion, 32 percent less, and 10 percent more. The average amount of promotion per pupil[1] was .74 semester. Probably the chief cause of the excess of demotions and failures and the low average promotion rate was the rather liberal promotion policy which had been pursued prior to the beginning of the project. Many decidedly inferior pupils were at the beginning of the experiment found to be almost up with normal pupils of the same chronological age, although they were unable to do satisfactory work as placed. Moreover, it had not been at all unusual to reward superior ability by allowing grades to be skipped.

The second half of Table V shows the changes made in the placement of the pupils in the control schools at this time. These changes were made by the teachers and principals according to the usual practice, which in Chapter I was called the "traditional method." A much larger percent of the pupils received normal promotion than in the experimental schools, but only about one-

[1]The average amount of promotion per pupil was computed by finding the total number of semesters of promotion given, subtracting therefrom the total number of semesters of demotion and dividing by the total number of pupils concerned.

half as many were demoted or failed or given extra promotion. The average amount of promotion per pupil was .87 semester.

It is not fair, however, to compare the promotion rates of the two groups of schools directly according to the figures given above. The general assumption as to the promotion rate is that it should be one semester per semester for pupils of normal mentality who are properly classified and working to their full capacity with no hindering factors entering into the situation. For pupils whose mentalities are above or below normal and who are working under the same conditions as those mentioned for normal pupils the theoretical rates of progress are proportionately above or below one semester per semester. For example, a pupil with an I. Q. of 125 would be expected to advance one and one-fourth semesters per semester and one with an I. Q. of 80, four-fifths of a semester per semester. Thus to render the average promotion figures given in the preceding paragraphs strictly comparable each should be divided by the average or median I. Q. of the pupils concerned in order to bring both to the basis of what they would be for pupils of normal mentality, that is, pupils whose I. Q. is 100.

The measure of progress obtained by dividing the actual average rate of progress per pupil by the median intelligence quotient of the pupils contributing to this average will be called the "progress quotient." It will be used as the true measure of progress throughout this study. Making use of this measure we have as the "progress quotient" of the experimental schools .74 divided by .94, and for the control schools .87 divided by 1.03. Thus the "progress quotients" at this time were 79 for the experimental schools and 84 for the control schools. The true difference in the promotion rate at the beginning of the experiment is thus seen to have been only .05 rather than .13 semester. This difference cannot be attributed to the operation of the experimental plan of organization but rather to conditions in the two groups of schools previous to the beginning of the experiment and to the preparation necessary before the project could be begun.

The formation of the fast, average and slow sections. As a result of the placement in February, 1921, more pupils were placed in the slow sections and fewer in the fast sections than would usually be the case in most school systems. This was largely due

to the liberal promotion policy that had been pursued before the experiment was begun, and to the fact that the pupil material of the experimental schools was rather distinctly inferior—median I. Q. 94, first quartile 80, third quartile 107. The fast sections included 14 percent of the total number of pupils placed, the average sections 41 percent, and the slow sections 45 percent.

The promotion and classification of the pupils for September, 1921. It was to be expected that after the experiment was under way a majority of the pupils in the experimental schools would make normal progress in the sections to which they had been assigned. The extent to which this expectation was fulfilled provided a measure of the efficiency of the previous placement. The supplementary principles of promotion and classification given in Chapter I suggest various reasons for the failure of many pupils to make such progress.

Table VI, which is similar to the first part of Table V, shows the gains and losses resulting from the placement of the pupils in the experimental schools for September, 1921. For example, in the average section of grade VIA 5 percent of the pupils were placed back in the VIA slow section and thus lost one-third of a semester; 17 percent placed in the VIIB slow section gained two-thirds of a semester; 63 percent placed in the VIIB average section gained one semester; 11 percent placed in the VIIB fast section gained one and one-half semesters and 4 percent placed in the VIIIA average section gained two semesters.

At this time 64 percent of the pupils of the experimental schools were advanced to the corresponding section of the next grade, but as such advancement meant only two-thirds of a semester for the slow pupils and one and one-half semesters for the fast, there were only 41 percent of the pupils who gained just one semester. In addition to the 64 percent mentioned, 26 percent were placed in the next grade but in a different section, so that in all 90 percent of the pupils of the experimental schools were advanced to the next grade. The average progress earned by the pupils in the slow sections was .65 semester, that earned by those in the average sections was .94 semester, and that by the members of the fast sections 1.38 semesters. For all the pupils the average was .88 semester. Dividing by the median I. Q.'s obtained from the Novem-

TABLE VI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1920-21

Grade	Section	Semesters Gained or Lost										
		−1⅓*	−1	−⅓	0	+⅓	+⅔	+1	+1⅓	+1⅔	+2	+2⅓
IB	Slow				14	22	60	226				.4
	Average				21		57				.4	
IA	Slow			4	4		46	43	2			1
	Average		1		7		12	58	18		1	1
	Fast			1			4	22	65		1	7
IIB	Slow			6	5		62	23	3			
	Average			7	4		18	69			1	
	Fast				3		7	24	62			3
IIA	Slow			5			70	25				
	Average		1	2	2	1	7	73	10		1	2
	Fast							20	76			4
IIIB	Slow			6	1		67	24	1			
	Average	1				1	22	58	17		1	
	Fast						2	32	60		2	4
IIIA	Slow			14	2		68	14	1			
	Average			5	3		21	63	6	1	1	
	Fast						1	42	51		4	1
IVB	Slow			10			68	22				
	Average			3	2		23	61	5			7
	Fast						.7	21	71			
IVA	Slow	2	1	7	6		60	23	1			
	Average	1		5			12	59	13		4	5
	Fast											
VB	Slow			6			63	30	2			
	Average			5			14	51	30			
	Fast							4	96			
VA	Slow	1		15			80	3	1			
	Average	1		6			17	52	24			
	Fast							16	84			
VIB	Slow			13			70	10	7			
	Average				2		25	54	16			2
	Fast							20	73			7
VIA	Slow			4			78	19				
	Average			5			17	63	11		4	
	Fast						5	20	75			
VIIB	Slow			2			76	20	2			
	Average			1			12	79	8			
	Fast								100			
VIIA	Slow				8		40	52				
	Average				1		4	76	19			
	Fast							6	94			
VIIIB	Slow			9			90	2				
	Average						4	96				
All	All	.2	.1	4	4	1	36	41	13	.1	1	1

*−1⅓ denotes a loss of one and one-third semesters, etc.

ber tests, the "progress quotients" were found to be 79, 92 and 118 for the slow, average and fast sections, respectively. For all the pupils of the experimental schools the quotient was 93.

Table VII shows the same data for the control schools as Table VI for the experimental. Of all the pupils in the control schools 80 percent were advanced just one semester. This percent is practically twice as large as that of the experimental schools, but is 10 smaller than the percent of pupils of those schools advanced to the same section of the next grade. Only 6 percent of the pupils of the control schools received extra promotion, as compared with the 15 percent in the experimental schools, but 14 percent were failed or demoted, as compared with only 8 percent in the latter group. The average progress in the control group was .92 semester. Dividing this by 1.02, a "progress quotient" of 90 was obtained. Thus, although the average progress of the pupils of the control schools was .04 semester greater, their "progress quotient" was three points smaller. To make the comparison upon a strictly valid basis, however, the effect of the pupils who left school and who entered school during the semester must be considered. Making the proper corrections for these pupils,[2] the average rate of progress

TABLE VII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1920–21

Grade	Semesters Gained or Lost							
	−2	−1	0	+1	+2	+3	+4	+5
IB			45	53	2			
IA		1	15	71	13	.4		
IIB			17	81	1			
IIA	1		10	88	1			
IIIB			12	87		.4	1	.4
IIIA		2	16	81	.4			
IVB			7	85	7	1		
IVA		2	9	88	1			
VB			11	87	2	1		
VA		1	4	83	12			
VIB			1	90	8	1		
VIA			9	75	16			
VIIB			1	90	9			
VIIA		.5	2	86	11			
VIIIB			4	96				
All	.1	1	13	80	6	.2	.1	.03

[2]See Appendix A.

of the pupils of the experimental schools was .02 semester less and their "progress quotient" four points larger than the corresponding figures for the pupils of the control schools. In other words, in so far as the progress of the pupils was concerned, the experimental schools were more efficient during the second semester of 1920-21 than were the control schools. The difference in the "progress quotient" was just about large enough to balance the difference at the beginning of the experiment.[8] Since this was the case, it cannot be assumed that the increased efficiency of the experimental schools in the matter of progress was necessarily due to the plan of organization used.

The relative size of the fast, average and slow sections. It was again deemed advisable to place many more pupils in the slow than in the fast sections. The facts that had made this necessary a semester earlier still exerted some influence upon the situation. The percent of the pupils placed in fast sections at this time was 15, and the remainder were equally divided between the average and the slow sections. Thus there was an increase of one percent in the number of pupils placed in the fast sections and also of those in the average sections over the percents for the previous semester.

The placement of the new entrants received by the experimental schools in September, 1921. The new entrants into the experimental schools in September, 1921, were tentatively placed in the average sections of the grades indicated by their previous school records, and later, after being tested, they were placed as the test results and the other data indicated. In the placement of pupils at this time no reclassification of the IB pupils was attempted, because the results obtained from the use of the Kingsbury Primary Group Intelligence Scale in that grade were so unsatisfactory that the writer deemed it unwise to make use of them. Table VIII shows that slightly over half of the pupils were not shifted at all, that 15 percent gained by the reclassification and 31 percent lost by it. The average change made amounted to a loss of .09 semester, or a promotion of .91 semester from the grades in which these pupils were the previous semester. As the median I. Q. of the new entrants was 85, their "progress quotient" was 107.

[8]See page 29.

TABLE VIII. PERCENTS OF THE NEW ENTRANTS INTO THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS BY THEIR RECLASSIFICATION IN SEPTEMBER, 1921

Temporary Grade	Semesters Gained or Lost											
	−2⅓	−1⅓	−1	−⅓	0	+½	+⅔	+1	+1½	+1⅔	+2	+2½
IB*												
IA			27	13	53			7				
IIB		10	2	14	52	11	2	10				
IIA		8		35	42	8		4	4			
IIIB		7		7	51	21		14				
IIIA				23	77							
IVB		10	2	17	58	6		6	2			
IVA				38	52	5					5	
VB		7		25	53	2		14				
VA		19		21	47	8	4	2				
VIB		9		14	59			9	5			5
VIA		17	13	17	43			4	4			
VIIB	3			12	64	15		3		3		
VIIA		29	7	7	50				7			
VIIIB		12	31	15	42							
VIIIA					100							
All	.2	9	4	18	53	7	.2	6	1	.2	.2	.2

*As is explained in the text, the pupils in grade 1B were not reclassified at this time.

Only 9 percent of the new entrants at this time were placed in the fast sections. The average sections received 63 percent and the slow sections 28 percent. Combining the new entrants with the pupils who had been tested in the previous May the percent in the fast sections was 15, that in the average 45, and that in the slow 41.

The promotion and classification of the pupils for February, 1922. The data obtained in December, 1921, were used to determine the placement of the pupils for the second semester of the school year. Table IX, which is similar to Table VI, gives the percents of the pupils of the experimental schools gaining or losing various amounts during the first semester of 1921-22. There were 48 percent of the pupils advanced just one semester, as compared with 41 percent during the previous semester; 35 percent made less than one semester's progress and 17 percent made more, as compared with 44 and 15 percent previously. The percent of the pupils advanced to the corresponding section of the next higher grade was 79, whereas only 64 percent were so advanced a semester previously. The percent placed in some section of the next grade

TABLE IX. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE FIRST SEMESTER OF 1921-22

Grade	Section	Semesters Gained or Lost																
		−1⅓	−1	−⅔	−½	−⅓	0	+⅓	+½	+⅔	+⅚	+1	+1⅓	+1½	+1⅔	+2	+2⅓	+2½
IB	Slow						7			61		27				5		
	Average						24			12		62				2		
IA	Slow						14	8		65			12					
	Average		1			3	6	7				76				7		
	Fast						2							56		41		
IIB	Slow					12		9		58			16		1		4	
	Average			2			4	4				86	2			2		
	Fast								3		1			95				1
IIA	Slow					12		8		73			8					
	Average	1		4				4				89				2		
	Fast								10		5			85				
IIIB	Slow					6				80			14					
	Average			1			5	3				84				7		
	Fast								2					98				
IIIA	Slow					6		2		82			10					
	Average	1		1			7	8				83						
	Fast								3					97				
IVB	Slow					4		2		79			16					
	Average			3			3	7				84				3		
	Fast										17			83				
IVA	Slow					1		1		76			22					
	Average			4				3				78	5			10		
	Fast													100				
VB	Slow					4				78			18					
	Average		1	5				4				85				5		
	Fast													100				
VA	Slow									92			7				1	
	Average			1			6	3				85	1			4		
	Fast								2					98				
VIB	Slow					1				63			36					
	Average						6	7				87						
	Fast				3									97				
VIA	Slow									59			37				4	
	Average	1	2	5			9	5				75				2		
	Fast													100				
VIIB	Slow					3				87			10					
	Average			1				4				94				1		
	Fast													100				
VIIA	Slow									74			26					
	Average			1			2					96						
	Fast													100				
VIIIB	Slow									83							17	
	Average			4			2	7				80				8		
	Fast													100				
VIIIA	Slow									100								
	Average							2				98						
All	All	.1	.1	1	.02	1	6	3	.2	23	.2	48	5	10	.02	2	.2	.02

was almost the same as before, being 89. The average progress per pupil was .72 semester for those in the slow sections, .87 semester for those in the average and 1.49 semesters for the members of the fast sections. The respective "progress quotients" were 84, 84 and 121. For all the pupils in the experimental schools the average progress was .90 semester and the "progress quotient" 91. The corresponding figures for the second semester of 1920-21 were .88 semester and 93, so it is apparent that the average progress was slightly greater and the "progress quotient" slightly less during the second semester of the experiment than during the first.

Table X, which is similar to Table VII, shows the gains and losses of the pupils of the control schools according to their placement at this time. A comparison of these data with those for the experimental schools shows that, as before, a larger percent of the pupils of the control schools received normal promotion. The difference, however, was not quite as great as the previous semester, the figures for this time being 82 and 48 percent as compared with 80 and 41 percent. The percent of the pupils receiving extra promotion in the control schools was only half as large as it had been a semester earlier, whereas in the experimental schools the corre-

TABLE X. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE FIRST SEMESTER OF 1921–22

Grade	Semesters Gained or Lost						
	−3	−2	−1	0	+1	+2	+3
IB				33	64	2	.2
IA				19	80	.5	
IIB				14	82	3	
IIA				17	83	1	
IIIB	.3		.3	6	91	3	
IIIA				20	80		
IVB				10	83	7	
IVA				6	92	2	
VB				10	84	4	2
VA			1	6	91	3	
VIB		1		11	85	4	
VIA	1			13	82	4	1
VIIB				11	88	1	
VIIA			1	8	79	13	
VIIIB				9	91		
VIIIA	1		2	9	89		
All	.1	.03	.1	14	82	3	.2

sponding percent was slightly larger. The percent of failures and demotions in each group was practically the same as before. The average progress per pupil in the control schools was .88 semester and the "progress quotient" was likewise 88. Hence the average progress was .02 semester greater in the case of the experimental schools and the "progress quotient" three points greater. Had it not been for the new entrants and eliminees, the difference in average progress would have been .01 semester greater. Thus it can be said for the second semester of the experiment, as for the first, that in so far as the progress of the pupils was concerned, the experimental schools were somewhat more efficient than were the control schools.

The classification into fast, average and slow sections for the second semester of 1921-22. The percents of the pupils placed in the sections at this time differed rather markedly from those for previous semesters. The percent placed in the fast sections showed only a slight decrease, but that in the slow sections decreased about one-third. The percents were 13 in the fast sections, 57 in the average and 30 in the slow sections. These figures give evidence that as the experiment progressed it was possible to place pupils more nearly as would be expected from theoretical considerations.

The promotion and classification of the pupils for September, 1922. After the testing in May, 1922, which was the last during the experiment, the pupils of the experimental schools were placed for the first semester of 1922-23. Table XI, which is similar to Tables VI and IX, shows the gains and losses of the pupils of the experimental schools during the second semester of 1921-22. There were 55 percent of the pupils who gained just one semester as compared with 48 percent during the previous semester, 28 percent who made less than one semester's progress as compared with 35 percent, and 17 percent who made more, the same as the previous semester. Only 58 percent of the pupils were advanced to the corresponding section of the next higher grade as compared with 79 percent a semester previously. The percent placed in some section of the next grade was 89, just the same as it had been. The average progress per pupil was .62 semester for those in the slow section, .98 semester for those in the average, and 1.39 for

TABLE XI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1921-22

Grade	Section	Semesters Gained or Lost													
		−1⅓	−1	−⅔	−⅓	0	+⅓	+½	+⅔	+1	+1⅓	+1⅔	+2	+2⅙	+2⅓
IB	Slow					37			46	17					
	Average					12			22	66					
IA	Slow			1		30	33			36					
	Average					20				69	11		.3		
	Fast					3			5	21	66		5		
IIB	Slow				10				90						
	Average					9				75	4		12		
	Fast									21	79				
IIA	Slow				5				94			2			
	Average					5				81	12		2		
	Fast									4	83		13		
IIIB	Slow				19				81						
	Average					12				81	7				
	Fast									16	65		18		
IIIA	Slow				11				89						
	Average					5				79	14		2		
	Fast										95				5
IVB	Slow				6				94						
	Average					5				77	18				
	Fast									19	78		4		
IVA	Slow				6				94						
	Average		1			6				82	8		3		
	Fast										100				
VB	Slow	1			3				96						
	Average					2				93	4		1		
	Fast										100				
VA	Slow								100						
	Average					3				86	7		3		
	Fast														
VIB	Slow								99					1	
	Average					2				85	6		6		
	Fast									8	92				
VIA	Slow								100						
	Average					5		6		83	6				
	Fast									18	82				
VIIB	Slow				3				26			71			
	Average					3				92	2		2		
	Fast									31	63		6		
VIIA	Slow				6				94						
	Average									89	8		3		
	Fast										100				
VIIIB	Slow				3				97						
	Average					4				96					
	Fast									38			63		
VIIIA	Slow									100					
	Average					4				96					
All	All	.03	.03	.03	1	6	1	.2	20	55	14	1	2	.03	.03

members of the fast sections. The respective "progress quotients" were 76, 98 and 114. For all pupils of the experimental schools the average progress was .94 semester and the "progress quotient" 97. These figures show both greater actual progress and greater progress relative to ability than was made during either of the previous semesters.

Table XII, which is similar to Tables VII and X, shows the gains and losses of the pupils of the control schools for this semester. Again more pupils of the control schools received normal promotion than was the case in the experimental schools, the difference, however, being smaller than it was in either of the previous semesters. The percent of the pupils receiving extra promotion was only one-third as large as in February, 1922, whereas in the experimental schools it was the same. The percent of failures and demotions in the control schools was slightly less than in February, the decrease being in about the same ratio as that in the experimental schools. The average progress per pupil in the control schools was .89 semester and the "progress quotient" 86. Thus the average progress was .05 semester greater in the case of the experimental schools and the "progress quotient" eleven points greater.

TABLE XII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE SECOND SEMESTER OF 1921–22

Grade	Semesters Gained or Lost				
	−1	0	+1	+2	+3
IB		8	69	23	
IA	.4	18	81	1	
IIB		16	83	2	
IIA		10	90		
IIIB		8	91	1	
IIIA		24	75	1	
IVB		8	90	2	
IVA		10	90		
VB		13	86	1	1
VA		6	94		
VIB		8	92		
VIA		7	92	1	
VIIB	3	9	85	1	1
VIIA		16	83	1	
VIIIB		17	83		
VIIIA		5	95		
All	.2	12	87	1	.1

The new entrants and eliminees did not affect these differences.

The classification into fast, average and slow sections for the first semester of 1922-23. The percents of the pupils placed in the sections for September, 1922, again differed considerably from those for previous semesters. The percent in the fast sections was practically the same, 14, but that in the average sections rose to 68 and that in the slow sections dropped to 18. Thus the tendency already noted for the fast and slow sections to approximate each other in size was continued as the experiment progressed longer. Probably the distribution at this time was about what it should be, as there will always be more pupils belonging in slow sections because of not realizing their highest possible achievement than there will be pupils belonging in fast sections because of doing more than should be expected of them.

Summary. As a result of the placement of the pupils at the beginning of the experiment the "progress quotient" for the experimental schools was, at that time, five points smaller than that for the control group. During the course of the experiment this situation was reversed. Averaging the "progress quotients" for the three semesters, those for the experimental group were the larger by about six points. Thus the net result of the experimental plan of organization in so far as progress was concerned was favorable. The greater degree of efficiency of the experimental schools seems to have been due to the operation of this plan.

CHAPTER IV

THE EFFICIENCY OF THE TWO GROUPS OF SCHOOLS AS MEASURED BY THE ACHIEVEMENTS OF THE PUPILS

In Chapter III, one of the two factors which were taken as constituting output has been discussed. In this chapter the other factor, that of achievement, will be considered both absolutely and in its relation to capacity. It has been shown in the preceding chapter that there was an increase in the "progress quotient" of the experimental schools as compared with that of the control schools. Therefore if a study of the achievements of the pupils of the two groups of schools shows that those of the experimental schools were either equal to or greater than those of the control schools, it may be said that the experimental schools were more efficient than the other group during this experiment.

The gains in absolute achievement during the second semester of 1920-21. Table XIII gives the median achievement ages of the grades and sections of the two groups of schools at the beginning and end of the second semester of 1920-21. A comparison of the first and fourth columns shows that at the beginning of this semester the median achievement age of the control schools was four months greater than that of the experimental schools, while at the end of this semester the median ages were the same. In other words, the gain in achievement age on the part of the experimental schools was four months more than that of the control schools. The cause of this increase cannot be stated with certainty. There are at least two explanations that may account for it. One of these is that it resulted from the same causes which accounted for a similar increase in the mental ages. The increase in the median mental age of the pupils of the experimental schools was five-tenths of a year greater during this semester than was that of the control schools. In the opinion of the writer the most potent cause of the greater increases in both mental and achievement ages on the part of the experimental schools was the fact that both the teachers and the pupils of those schools felt a very high degree of

TABLE XIII. MEDIAN GRADE AND SECTION ACHIEVEMENT AGES OF THE PUPILS OF THE TWO GROUPS OF SCHOOLS AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Grade	Section	Experimental			Control		
		Beginning	End	Gain*	Beginning	End	Gain
IA	Slow	6–4	10–5	49			
	Average	6–8	10–4	44	6–2	10–0	46
	Fast	5–6	10–10	64			
IIB	Slow	6–4	8–8	28			
	Average	6–5	9–10	41	7–11	8–0	1
	Fast	8–6	8–4	–2			
IIA	Slow	6–7	9–5	34			
	Average	7–7	9–10	27	7–10	9–7	21
	Fast	8–1	9–6	17			
IIIB	Slow	8–2	10–10	32			
	Average	8–11	10–7	20	9–0	10–9	21
	Fast	8–8	11–1	29			
IIIA	Slow	9–1	9–1	0			
	Average	10–1	10–1	0	9–8	10–1	5
	Fast	9–10	9–8	–2			
IVB	Slow	8–11	10–8	21			
	Average	10–0	10–7	7	9–1	10–10	21
	Fast	9–11	12–6	31			
IVA	Slow	9–5	11–7	26			
	Average	10–7	14–0	41	10–0	10–0	0
	Fast	....	....	...			
VB	Slow	9–8	12–1	29			
	Average	10–10	13–4	30	10–7	12–2	19
	Fast	10–10	13–2	28			
VA	Slow	10–4	11–4	12			
	Average	11–5	13–6	25	11–5	13–4	23
	Fast	13–0	14–7	19			
VIB	Slow	10–0	12–1	25			
	Average	11–10	12–10	12	12–0	13–8	20
	Fast	13–6	15–1	19			
VIA	Slow	11–11	14–5	30			
	Average	12–8	15–1	29	12–1	13–10	21
	Fast	15–0	15–0	0			
VIIB	Slow	12–5	15–5	36			
	Average	12–5	16–6	49	14–2	14–8	6
	Fast	15–4	17–0	20			
VIIA	Slow	12–7	14–11	28			
	Average	13–6	17–8	50	15–0	16–1	13
	Fast	16–1	18–5	28			
VIIIB	Slow	11–11	13–6	19			
	Average	14–0	16–7	31	15–1	16–1	12
All	All	10–0	11–6	18	10–4	11–6	14

*The gains are given in terms of months.

TABLE XIV. MEDIAN GRADE AND SECTION ACHIEVEMENT QUOTIENTS OF THE TWO GROUPS OF SCHOOLS AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Grade	Section	Experimental			Control		
		Beginning	End	Gain	Beginning	End	Gain
IA	Slow	103	116	13			
	Average	88	118	30	91	117	26
	Fast	65	128	63			
IIB	Slow	93	100	7			
	Average	86	109	23	99	95	−4
	Fast	103	105	2			
IIA	Slow	89	112	23			
	Average	92	111	19	90	111	21
	Fast	88	110	22			
IIIB	Slow	111	120	9			
	Average	106	111	5	103	118	15
	Fast	103	112	9			
IIIA	Slow	114	104	−10			
	Average	114	115	1	115	112	−3
	Fast	108	104	−4			
IVB	Slow	111	106	−5			
	Average	106	106	−0	110	112	2
	Fast	106	108	2			
IVA	Slow	107	110	3			
	Average	106	118	12	103	103	0
	Fast	...	...	...			
VB	Slow	109	112	3			
	Average	108	109	1	105	110	5
	Fast	98	108	10			
VA	Slow	106	108	2			
	Average	110	116	6	106	112	6
	Fast	105	118	13			
VIB	Slow	102	110	8			
	Average	104	108	4	109	114	5
	Fast	110	106	−4			
VIA	Slow	115	119	4			
	Average	108	118	10	105	115	10
	Fast	115	121	6			
VIIB	Slow	108	120	12			
	Average	105	122	17	103	108	5
	Fast	105	117	12			
VIIA	Slow	108	106	−2			
	Average	103	119	16	101	113	12
	Fast	105	122	17			
VIIIB	Slow	99	104	5			
	Average	93	113	20	105	111	6
All	All	106	112	6	104	112	8

interest in the results of the tests because they knew that placement was largely dependent upon these results. On the other hand, the teachers and pupils of the control schools knew that no direct use would be made of the test results, hence naturally took less interest

in the testing. The other explanation is that the experimental plan of organization caused the increase. Inasmuch as there is generally a fairly high correlation between the scores made on intelligence tests and those on subject-matter tests, especially in the case of verbal intelligence and reading tests, the writer believes that the first explanation is the true one or at least more nearly so than the latter. It is not unlikely that both had a part in causing the relative increase.

The achievement quotients at the beginning and end of the second semester of 1920-21. Table XIV shows that at the beginning of this semester the median achievement quotient of the experimental schools was two points higher than that of the control schools. We have seen that both the intelligence and the achievement scores made at the end of the semester showed a greater increase in the case of the experimental schools than in that of the other group, but that the increase in intelligence on the part of the experimental schools was slightly greater than that in achievement. Moreover, both groups showed greater increases in achievement than in intelligence. Therefore we expect to find, as we do, that the median achievement quotients of both groups of schools increased during this semester, and that the increase in the case of the control schools was slightly greater. This difference was two points. Thus at the end of the semester the two medians were the same. The general import of this evidence is that in so far as achievement was concerned there was a slight relative increase in the efficiency of the control schools.

The correlation of intelligence and achievement at the beginning and end of the second semester of 1920-21. Although the achievement quotient measures the relation of intelligence and achievement in one way, this relation may also be measured by means of the coefficient of correlation. The following table shows the coefficients that were found at the beginning and end of the semester by correlating the mental and achievement ages for all grades combined:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
$.68 \pm .01$	$.56 \pm .01$	.12	$.60 \pm .01$	$.51 \pm .01$	.09

It is evident that in both groups there was a decrease in the

correlation of achievement with intelligence as measured by the tests used. This decrease was slightly greater in the case of the experimental schools, but the difference was not great enough to be significant. It may be that this decrease was due to a lessening of the degree to which instruction was adapted to the capacities of the pupils. It is likely, however, that much, if not all, of the decrease was caused by the fact that the mental ages calculated at the beginning of the semester were based upon the average scores made on two intelligence tests and hence were more reliable than those obtained at the end of the semester, which were based upon only one test score. This lower degree of reliability would naturally tend to reduce the correlation between the mental and the achievement ages.

The relation of the intelligence and achievement quotients at the beginning and end of the second semester of 1920-21. The median achievement quotients of the groups of different levels of intelligence as determined by the intelligence quotients were computed for the end of the semester as they had been at its beginning. Table XV presents a comparison of those found at the two periods. The achievement quotients of the different groups at the end of the semester showed that in the experimental schools instruction was adapted about equally well to the pupils of different

TABLE XV. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1920–21

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150–59	104	117	98	103
140–	95	122	95	113
130–	102	113	98	107
120–	102	113	102	109
110–	101	112	101	109
100–	102	113	102	113
90–	104	111	104	113
80–	103	113	105	114
70–	107	115	110	117
60–	113	116	114	120
50–	109	128	125	120
All	105	112	104	112

levels of intelligence, whereas in the control schools the previous well-marked tendency to adapt instruction more nearly to the capacities of the inferior pupils remained. As this tendency was present in both groups of schools at the beginning of the semester it is evident that there was a relative improvement in the degree to which the instruction in the experimental schools was adapted to pupils of one level of intelligence as well as to those of another.

The coefficients of correlation of the intelligence and achievement quotients were also found and compared with those for the beginning of the semester. The following table presents this comparison:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Gain
—.16±.01	—.12±.01	.04	—.28±.01	—.25±.01	.03

This comparison shows that at the close of the semester the negative correlations between the intelligence and achievement quotients were slightly smaller in the cases of both of the groups of schools. The difference in the gains was so small that it has no significance. This fact shows that the instruction given in the control schools was still somewhat less equally suited to pupils of all levels of intelligence than was that of the experimental schools. In the main this corroborates the evidence presented in the preceding paragraph.

The gains in absolute achievement during the first semester of 1921-22. Table XVI, which contains data corresponding to the third and sixth columns of Tables XIII and XIV, shows that the increase in absolute achievement during this semester was seven months of achievement age in the experimental schools and four months in the control schools. As during this same semester the median mental age of the experimental schools did not increase as much as did that of the control schools it seems fair to attribute the greater gain in achievement to an increase in the efficiency of the experimental schools. It was shown in the preceding chapter that during this same semester the "progress quotient" of this group of schools was greater than that of the control group, therefore this increase in efficiency in so far as achievement is concerned cannot be attributed to a slowing up of the progress of the pupils.

TABLE XVI. GRADE AND SECTION GAINS AND LOSSES IN ACHIEVEMENT AGES AND QUOTIENTS OF THE TWO GROUPS OF SCHOOLS FROM THE BEGINNING TO THE END OF THE FIRST SEMESTER OF 1921–22

Grade	Section	Achievement Ages		Achievement Quotients	
		Experimental	Control	Experimental	Control
IA	Slow	16		12	
	Average	23	−2	7	6
	Fast	8		−19	
IIB	Slow	−20		−11	
	Average	−14	−18	−19	−31
	Fast	−11		−13	
IIA	Slow	11		18	
	Average	7	6	−2	−21
	Fast	5		10	
IIIB	Slow	8		−1	
	Average	17	12	−11	−32
	Fast	16		−8	
IIIA	Slow	−15		−11	
	Average	−17	−16	−7	−12
	Fast	−10		−10	
IVB	Slow	10		−4	
	Average	15	4	−4	−9
	Fast	21		−5	
IVA	Slow	23		1	
	Average	22	−7	1	−18
	Fast	67		12	
VB	Slow	20		−11	
	Average	5	17	−10	−1
	Fast	−2		−25	
VA	Slow	12		−7	
	Average	−6	18	−15	2
	Fast	6		−9	
VIB	Slow	17		−8	
	Average	19	8	−3	−8
	Fast	−49		−12	
VIA	Slow	14		−13	
	Average	2	8	−11	−5
	Fast	−29		−2	
VIIB	Slow	19		−9	
	Average	18	0	−10	−12
	Fast	35		2	
VIIA	Slow	5		−13	
	Average	10	3	−3	−5
	Fast	27		−3	
VIIIB	Slow	−13		−15	
	Average	−8	1	−13	−9
	Fast	−15		−18	
VIIIA	Slow	7		−7	
	Average	12	−3	3	−17
All	All	7	4	−8	−13

The achievement quotients at the beginning and end of the first semester of 1921-22. The median achievement quotients of both groups of schools were smaller at the end of this semester than they were at its beginning. In other words, the average increase in the scores made upon the intelligence tests was considerably greater than that in those upon the achievement tests. This would seem to point to the fact that the practice effect upon the intelligence tests was greater than that upon the others. As Table XVI shows, the decrease in the median achievement quotient of the experimental schools was eight points, whereas that in the control schools was thirteen points. Thus the loss of the experimental schools was five points less than that of the other group, or, in other words, their relative gain was that large. This supports the conclusion reached above from the study of the mental and achievement ages, that the efficiency of the experimental schools during the semester was greater than that of the control schools.

The correlation of achievement and intelligence at the beginning and end of the first semester of 1921-22. The following table compares the coefficients of correlation found at the end of the semester with those at the beginning:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
.56±.01	.54±.01	.02	.51±.01	.51±.01	.00

Judging from these coefficients, it seems that there was practically no change in the relation of achievement to intelligence during this semester. The slight decrease of .02 on the part of the experimental schools was too small to have any significance.

The relation of the intelligence and achievement quotients at the beginning and the end of the first semester of 1921-22. Table XVII shows the same facts for this semester as Table XV for the previous semester. At the end of this semester there was a rather definite decrease in the achievement quotient medians of the experimental schools from the duller to the brighter pupils. This tendency was even more marked in the control schools. Inasmuch as at the beginning of the semester this tendency was not noticeable in the experimental schools but was present in the control schools the figures for the end of the semester indicate that relatively the condition which they measure became worse in the

TABLE XVII. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE FIRST SEMESTER OF 1921-22

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150-59	117	98	103	85
140-	122	96	113	88
130-	113	102	107	95
120-	113	100	109	99
110-	112	102	109	104
100-	113	106	113	105
90-	111	106	113	105
80-	113	110	114	107
70-	115	111	117	113
60-	116	114	120	119
50-	128	125	120	132
All	112	105	112	104

experimental schools. That is to say, during this semester there was a relative loss in the degree to which instruction was equally well adapted to pupils of all levels of intelligence in the experimental schools. The writer is unable to suggest any probable explanation of this fact.

A comparison of the coefficients of correlation of the intelligence and achievement quotients at the end of the semester with those at the beginning supports the conclusion given above. These coefficients were as follows:

Experimental			Control		
Beginning	End	Loss	Beginning	End	Loss
—.12±.01	—.36±.01	.24	—.25±.01	—.39±.01	.14

This comparison shows that the correlation between the intelligence and achievement quotients became considerably greater, negatively, during the semester. The change was much larger in the experimental schools. This fact emphasizes the conclusions presented in the last two paragraphs to the effect that there was a relative decrease in the degree to which the experimental schools capitalized the capacities of their pupils into achievement. This decrease is even more definitely shown by these coefficients than by the data given previously.

The gains in absolute achievement during the second semester of 1921-22. Table XVIII, which is similar to Table XVI,

TABLE XVIII. GRADE AND SECTION GAINS AND LOSSES IN ACHIEVEMENT AGES AND QUOTIENTS OF THE TWO GROUPS OF SCHOOLS FROM THE BEGINNING TO THE END OF THE SECOND SEMESTER OF 1921–22

Grade	Section	Achievement Ages		Achievement Quotients	
		Experimental	Control	Experimental	Control
IA	Slow Average Fast	35 40 47	18	25 45 16	4
IIB	Slow Average Fast	3 7 −7	0	7 12 5	−4
IIA	Slow Average Fast	11 10 6	15	7 10 25	17
IIIB	Slow Average Fast	1 7 7	16	−3 2 10	19
IIIA	Slow Average Fast	−16 −15 −2	−4	−6 −9 16	13
IVB	Slow Average Fast	6 4 15	16	1 9 3	12
IVA	Slow Average Fast	5 13 12	19	7 9 20	11
VB	Slow Average Fast	0 10 7	12	8 11 11	11
VA	Slow Average Fast	8 12 10	3	13 10 13	4
VIB	Slow Average Fast	20 17 16	5	23 13 17	5
VIA	Slow Average Fast	6 5 4	3	4 8 3	7
VIIB	Slow Average Fast	34 9 −9	−12	17 5 10	−1
VIIA	Slow Average Fast	−23 2 9	6	−5 8 9	2
VIIIB	Slow Average Fast	18 −6 −19	25	14 1 −1	17
VIIIA	Slow Average	6 −11	20	−16 1	11
All	All	7	4	7	6

shows that the increase in absolute achievement during the second semester of 1921-22 averaged seven months in the experimental schools and four months in the control schools. Thus again it appears that the experimental schools were more efficient as regards the achievement of their pupils. As their "progress quotient" was considerably greater during this semester the gain in absolute achievement can not be attributed to holding back the pupils.

The achievement quotients at the beginning and end of the second semester of 1921-22. Table XVIII likewise presents the gains of the two groups of schools in achievement quotients. According to these quotients the gain of the experimental schools was only one point greater than that of the control schools.

The correlation of achievement and intelligence at the beginning and end of the second semester of 1921-22. The coefficients of correlation between absolute achievement and intelligence at the beginning and end of this semester were as follows:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Gain
.54±.01	.75±.01	.21	.51±.01	.53±.01	.02

Judging from these coefficients it seems that there was a very decided gain in the relation of achievement to intelligence on the part of the experimental schools, but practically no change in the control schools.

The relation of the intelligence and achievement quotients at the beginning and end of the second semester of 1921-22. Table XIX, which is similar to Tables XIII and XVII, presents the relation of the achievement and intelligence quotients for the second semester of 1921-22. Comparing the figures for the beginning and end of this semester there seems to have been no noticeable change in the situation.

A comparison of the coefficients of correlation of the intelligence and achievement quotients is more favorable to the experimental schools. These coefficients were as follows:

Experimental			Control		
Beginning	End	Gain	Beginning	End	Loss
—.36±.01	—.34±.01	.02	—.39±.01	—.52±.01	.13

The change in the coefficients of the control schools showed a decrease in the adaptation of instruction to pupils of all levels of

TABLE XIX. MEDIAN ACHIEVEMENT QUOTIENTS OF THE PUPILS OF DIFFERENT LEVELS OF INTELLIGENCE AT THE BEGINNING AND END OF THE SECOND SEMESTER OF 1921–22

Intelligence Quotient	Achievement Quotients			
	Experimental		Control	
	Beginning	End	Beginning	End
150–59	98	89	85	88
140–	96	99	88	92
130–	102	106	95	101
120–	100	109	99	104
110–	102	108	104	109
100–	106	112	105	111
90–	106	115	105	112
80–	110	114	107	118
70–	111	121	113	119
60–	114	131	119	128
50–	125	125	132	131
All	105	113	104	109

intelligence, but in the experimental schools such adaptation seems to have remained about the same.

The achievements of the two groups of schools during the second semester of 1921-22 as measured by the Omnibus Test. In planning this experiment it was decided to make use of tests in reading and arithmetic because those are generally considered the two most important subjects of the elementary curriculum and further because it was believed that the results obtained would give a fairly reliable index of the achievements of the pupils in all subjects. In order to provide a partial check upon this latter assumption a test was devised by the writer and given to the pupils of grade VIB and above at the regular testing period in May, 1922. This test, which was called the Omnibus Test,[1] contained questions in geography, history, grammar, elementary science and certain phases of arithmetic not covered by the standardized tests used. The scores made on this test were translated into achievement ages and quotients in the same way as for the other tests of achievement.

Table XX presents the median ages and quotients for the various grades and sections of the two groups of schools. It may be seen from this table that the showing made upon this test by

[1]See Appendix C.

TABLE XX. MEDIAN GRADE AND SECTION OMNIBUS ACHIEVEMENT AGES AND QUOTIENTS OF THE PUPILS OF THE TWO GROUPS OF SCHOOLS AT THE END OF THE SECOND SEMESTER OF 1921–22

Grade	Section	Ages		Quotients	
		Experimental	Control	Experimental	Control
VIB	Slow	10–8		97	
	Average	12–3	10–10	102	88
	Fast	10–8		90	
VIA	Slow	10–8		85	
	Average	11–1	15–2	89	120
	Fast	12–2		85	
VIIB	Slow	11–10		95	
	Average	12–0	14–10	92	109
	Fast	15–3		100	
VIIA	Slow	11–1		86	
	Average	14–9	14–4	102	107
	Fast	16–8		114	
VIIIB	Slow	12–11		98	
	Average	13–5	15–7	87	108
	Fast	15–2		90	
VIIIA	Slow	11–7		88	
	Average	14–11	17–3	96	111
	All	12–8	14–11	94	107

the control schools was very much better than that made by the experimental schools. The average difference was over two years of achievement age and thirteen points of achievement quotient. Inasmuch as the pupils were not given a similar test at any previous time the relative gain can not be computed. The difference between the two groups is so great, however, that it is evidently significant. A very probable conclusion is that in the experimental schools there was a tendency to emphasize the instruction in reading and arithmetic to the neglect of that in the other subjects. This tendency was probably not due to the fact that the teachers and pupils were consciously striving to prepare to make better scores upon the tests but that merely through the use of the tests in reading and arithmetic attention was called to pupils' weaknesses in these subjects and therefore unusual care was taken to correct these weaknesses.

Individual opinion as to the relative importance of arithmetic

and reading as compared with the elementary school subjects covered by the Omnibus Test will largely determine one's belief as to whether or not the experimental schools made a relative gain in achievement during the course of the experiment. Inasmuch as there was an average relative gain of only about one point per semester in the achievement quotient on the part of the experimental schools it is the opinion of the writer that there was not any greater efficiency in the achievement of this group of schools. On the other hand, he does not believe it should be asserted that in so far as achievement was concerned there was a decidedly smaller degree of efficiency.

Summary. The data presented in this chapter considering them from the standpoint of the experimental schools relative to the control schools may be briefly summarized as follows:

1. At the beginning of the experiment the median achievement age as measured by the tests used was four months lower. During the experiment slightly greater efficiency was shown, averaging about one month per semester, according to the reading and arithmetic test results. According to the results on the Omnibus Test, however, the median achievement age was twenty-seven months lower at the conclusion of the experiment.
2. The median achievement quotient derived from the arithmetic and reading tests was two points greater both at the beginning and end of the experiment. Allowing for the effect of the new entrants and eliminees, however, there was a relative gain of about one point per semester. The Omnibus Test achievement quotient was thirteen points lower.
3. The correlation of intelligence and achievement was .08 greater at the beginning and .25 greater at the close of the experiment.
4. At the beginning of the experiment instruction was somewhat better adapted to the inferior than to the superior pupils in both groups of schools. On the whole there was little change in this situation.

Considering these items together it seems that in so far as

achievement was concerned the efficiency of the experimental schools was no greater than that of the control schools. The slightly greater efficiency in reading and arithmetic was at least balanced by the results of the Omnibus Test. If we assume that the measurement of achievement shows no advantage for either group of schools it may be said that the experimental plan of organization was more efficient than the traditional plan because of the fact that the progress of the pupils was considerably greater in the experimental than in the control schools. If, on the other hand, it is considered that the Omnibus Test showed a distinctly greater degree of efficiency as regards the total achievement for the control schools, this must be balanced against the greater progress made in the other group and a less definite conclusion reached.

CHAPTER V

A STUDY OF THE PUPILS WHO REMAINED IN SCHOOL THROUGHOUT THE COURSE OF THE EXPERIMENT

Although there was no reason to suspect that a study of the records of the pupils who participated in this experiment throughout the three semesters that it continued would yield results and conclusions materially different from those presented in Chapters III and IV, yet it was thought desirable to make such a study. Therefore this chapter will present certain data concerning the pupils who were in the schools in February, 1921, and remained therein until the close of the experiment. These pupils did not compose as large a group as might be expected because the pupil population of both groups of schools was very unstable. Slightly less than 60 percent of the pupils tested at the beginning of the experiment were still in the schools at its conclusion. In making a study of these pupils the tabulations were not made by separate semesters but all three semesters were taken together.

The placement of the pupils and their progress throughout the grades. Tables XXI and XXII show the percents of pupils in the two groups of schools gaining or losing the given number of semesters during the three semesters that the experiment continued. From these tables it may be seen that only 34 percent of the pupils of the experimental schools made just three semesters of progress, whereas 62 percent of those of the control schools did so. The percents making more than this amount of progress were 26 and 7, respectively, and those making less, 40 and 32. The average amount of progress made was 2.79 semesters in the experimental schools but only 2.67 semesters in the control schools. Dividing these figures by three to reduce them to a semester basis and then by the median I. Q.'s gives "progress quotients" of 93 and 89, respectively. Therefore it appears that in so far as progress was concerned the experimental plan of organization was somewhat more efficient for those pupils remaining in school throughout the experiment than was the traditional plan. The

TABLE XXI. PERCENTS OF THE PUPILS OF THE EXPERIMENTAL SCHOOLS PRESENT THROUGHOUT THE EXPERIMENT THAT GAINED OR LOST THE GIVEN NUMBER OF SEMESTERS

Feb., 1921		Semesters Gained or Lost															
Grade	Section	−5	−1	0	+1	+2	+2½	+3	+3½	+4	+4½	+5	+5½	+6	+7	+10	+11
IB	Slow				22	52	5	20		2							
	Average				11	28	1	33		25			2				
IA	Slow		1	3	9	38	1	38	8	2							
	Average				4	15		44	14	17		3			1	1	
	Fast				4			22	16	27	27	2	2				
IIB	Slow				23	51	2	15	8	2							
	Average				6	33	9	39	9	3							
	Fast					9	5	36		23	27						
IIA	Slow			1	8	59		25	4	1	1						
	Average			1	2	12		52	11	9	12						
	Fast					6		13	29	19	32						
IIIB	Slow			1	8	59		19	12			1					
	Average				6	27		47	8	2		4		4			2
	Fast				3	6	6	34	9	17	17	6	3				
IIIA	Slow			1	17	47		23	7	2	3						
	Average				7	20		39	16	10	4	3					
	Fast				2	3		32	13	22	22	2	5				
IVB	Slow				7	46		31	7	7		1					
	Average					19		36	8	25	6			6			
	Fast					2		1	1	3	2			1			
IVA	Slow			1	8	44		42	4								
	Average				2	16		41	2	20	11	5	2				
	Fast																
VB	Slow	1			5	54		26	5	5	2						
	Average					14		39	11	7	27	2					
	Fast					5		11		16	67						
VA	Slow				10	54		28	1	6							
	Average			2		22		44	13	11	8						
	Fast							20		60	13	7					
VIB	Slow				13	20		55	3	5	5						
	Average				3	14		57		12	12	2					
	Fast							13	3	7	77						
VIA	Slow			2	9	61		20	7								
	Average					20		46	18	16							
	Fast					10		10		80							
VIIB	Slow				7	60		33									
	Average					16	2	83									
	Fast																
VIIA	Slow					100											
	Average				8	92											
	Fast					100											
VIIIB	Slow																
	Average				100												
VIIIA	Slow																
	Average																
All	All	.04	.04	.4	7	32	1	34	7	10	7	1	.4	.2	.04	.04	.04

TABLE XXII. PERCENTS OF THE PUPILS OF THE CONTROL SCHOOLS PRESENT THROUGHOUT THE EXPERIMENT THAT GAINED OR LOST THE GIVEN NUMBER OF SEMESTERS

February 1921 Grade	−1	0	+1	+2	+3	+4	+5	+6
IB		1	16	51	30	3		
IA		1	7	24	55	13		
IIB			7	37	56	1		
IIA		1	4	31	60	3		
IIIB		1	6	25	64	2		3
IIIA		4	5	26	58	7		
IVB			3	16	76	3	2	
IVA			1	23	73	2		
VB			1	17	78	2		1
VA			3	17	62	17	1	
VIB				21	69	10		
VIA			2	22	64	12		
VIIB			5	20	75			
VIIA	7		7	87				
VIIIB								
All	.1	1	5	26	62	6	.2	.3

difference of four points in the "progress quotients," especially when it is remembered that this is an average difference for three semesters, is large enough to justify the above statement.

The achievements of the pupils. Table XXIII presents the median achievement ages and quotients for the two groups of schools at the beginning and end of the experiment. From these data it may be seen that the gain on the part of the pupils of the experimental schools was three months of achievement age greater than that for the other group of schools and that the gain in achievement quotient was one point greater. These figures show that for the pupils who remained throughout the course of the experiment the experimental schools were slightly more efficient in so far as achievement was concerned.

Summary. The evidence afforded by the study of the pupils who remained in school during the course of the experiment shows that for these pupils the experimental plan of procedure resulted in appreciably greater progress according to the ability of the children and in slightly greater achievement. The difference in the "progress quotients" was four points and that in the achievement quotients, one point. Thus the general conclusion to be drawn

TABLE XXIII. MEDIAN ACHIEVEMENT AGES AND QUOTIENTS OF PUPILS WHO WERE PRESENT THROUGHOUT THE EXPERIMENT IN FEBRUARY, 1921 AND MAY, 1922

1921		Achievement Age				Achievement Quotient			
		Experimental		Control		Experimental		Control	
Grade	Section	1921	1922	1921	1922	1921	1922	1921	1922
IA	Slow	6–5	7–7			101	113		
	Average	5–10	9–7	6–6	9–2	88	132	91	108
	Fast	5–6	11–6			70	125		
IIB	Slow	6–6	6–5			100	100		
	Average	6–7	8–4	7–7	7–6	90	106	101	93
	Fast	8–0	9–5			103	108		
IIA	Slow	6–7	8–11			90	112		
	Average	7–6	9–8	8–1	9–6	93	111	93	105
	Fast	8–1	9–6			91	102		
IIIB	Slow	8–2	9–5			109	118		
	Average	8–10	10–0	9–1	10–8	106	107	105	104
	Fast	8–10	8–11			98	109		
IIIA	Slow	8–11	8–8			115	105		
	Average	10–1	9–5	9–8	10–0	113	99	115	98
	Fast	9–10	10–6			107	92		
IVB	Slow	8–11	9–4			112	103		
	Average	9–11	10–4	9–1	10–11	106	110	108	115
	Fast	10–6	10–2			105	100		
IVA	Slow	9–7	10–7			107	111		
	Average	10–11	11–10	10–3	12–3	111	117	102	124
	Fast	10–6	12–10			100	111		
VB	Slow	9–7	11–8			107	116		
	Average	11–0	12–8	10–7	11–11	110	115	104	114
	Fast	10–5	12–10			98	114		
VA	Slow	10–2	12–0			106	117		
	Average	11–8	13–8	11–4	12–1	108	118	107	111
	Fast	11–11	16–10			103	137		
VIB	Slow	10–1	12–10			104	124		
	Average	11–11	14–4	12–1	14–1	106	118	107	120
	Fast	13–10	16–4			112	116		
VIA	Slow	11–7	13–2			112	111		
	Average	12–6	14–0	12–1	14–5	104	114	107	112
	Fast	15–0	14–6			112	110		
VIIB	Slow	12–5	15–2			108	123		
	Average	12–4	13–7	13–11	14–4	103	117	105	108
	Fast	14–6	15–10			103	118		
VIIA	Slow	12–6	13–7			110	110		
	Average	13–7	16–1	14–7	14–9	103	114	101	113
	Fast	15–6	18–2			102	124		
VIIIB	Slow	....	16–5			....	125		
	Average	12–0	17–0	16–5	16–7	100	114	118	120
	Fast		19–8			....	120		
VIIIA	Slow	....	14–0			....	95		
	Average	....	17–10	....	17–4	....	113	....	112
All	All	9–10	11–6	10–2	11–7	106	112	104	109

from this study is the same as that drawn from the study of all the pupils, that the experimental plan of organization was somewhat more efficient than was the traditional plan.

CHAPTER VI

A SPECIAL STUDY OF THE BRIGHTER AND DULLER PUPILS

It is evident that such an experiment as the one described in this bulletin might not have the same effect upon the efficiency of the instruction of the brighter, the average and the duller pupils. In view of this fact a special study was made of the brighter and another of the duller pupils in order to discover the effect of the experimental plan of organization upon the efficiency of the instruction of these two groups. For the purpose of the two studies the records of those pupils whose I. Q.'s as found at the first testing period were 115 or higher and of those whose I. Q.'s were less than 80 were used. The former group included about one-sixth of the total number of pupils and the latter group about one-fifth. All records not complete for the duration of the experiment were rejected so that the number of pupils actually included in these studies was reduced to 199 brighter pupils and 514 duller pupils from the experimental schools and 396 brighter and 291 duller pupils from the control schools.

The placement of the brighter pupils and their progress through the grades. Of the 199 pupils of the experimental schools 2 percent were placed in the slow sections, 23 percent in the average sections and 75 percent in the fast sections at the beginning of the experiment. When it closed the respective percents were 1, 51 and 49. The marked reduction of the number in the fast sections was due to the fact that by the close of the experiment these pupils had gained one semester or more and in many cases were not quite bright enough to attempt to make further gain, at least immediately.

Table XXIV shows that the number of semesters gained by the brighter pupils of the experimental schools varied from two to six, and by those of the control schools from one to four. The percents of the brighter pupils of the experimental schools making less than regular, regular and more than regular progress, were 9, 23, and 68, respectively. In the control schools the corresponding

TABLE XXIV. PERCENTS OF THE BRIGHTER AND OF THE DULLER PUPILS OF THE TWO GROUPS OF SCHOOLS GAINING OR LOSING THE GIVEN NUMBER OF SEMESTERS DURING THE EXPERIMENT.

	−5	0	1	2	2½	3	3½	4	4½	5	5½	6	7	10	11	Average Progress	"Progress Quotient"
Brighter Pupils																	
Experimental																	
Slow.......						25	25	50								3.63	97
Average....				13		42	22	11	7	2	2					3.29	92
Fast.......				5	2	17	9	19	40	3	3	2				3.97	106
All........				7	2	23	13	18	32	3	2	2				3.80	102
Control......			2	15		72		11								2.92	77
Duller Pupils																	
Experimental																	
Slow.......	.2	2	14	57	.2	20	4	3	.5	.2						2.14	99
Average....			13	27	2	43	3	8	13			1*	1*	1*	1*	2.91	133
Fast.......			13	25			13	38	1							3.13	149
All........	.2	1	14	54	.4	23	4	4		.2		.2*	.2*	.2*	.2*	2.26	105
Control......		1	8	36		50		5		.3		1*				2.56	119

*These large amounts of progress were made by foreign-born pupils who, at the beginning of the experiment, were so handicapped by their inability to use the English language that they made low test scores and did poor school work. Many of these pupils were able to skip the work of several semesters as soon as the language difficulty was overcome.

figures were 17, 72 and 11. Table XXIV also shows that the average progress of the brighter pupils of the experimental schools was .88 of a semester greater than that of the pupils of the control schools and that their "progress quotient" was twenty-five points greater. These differences show that the experimental schools were much more effective in so far as the rate of progress of the brighter pupils was concerned.

The achievements of the brighter pupils. The table just below gives the median achievement ages and quotients of the brighter pupils of both groups of schools in February, 1921, and May, 1922.

	Experimental			Control		
	1921	1922	Gain	1921	1922	Gain
Achievement Age	11–4	12–11	19	11–2	12–7	17
Achievement Quotient	103	112	9	98	104	6

From these data it is evident that the more rapid progress of the pupils of the experimental schools did not result in a lessening of their relative achievement but was accompanied by a small gain. This gain in relative achievement amounted to two months in terms of achievement age or three points in terms of achievement quotient. Thus considering progress and achievement together, it may be said that for the brighter pupils the experimental plan of organization resulted in a marked increase of efficiency.

The placement of the duller pupils and their progress through the grades. Of the 514 duller pupils from the experimental schools, 86 percent were placed in the slow sections, 12 percent in the average and 2 percent in the fast sections at the beginning of the experiment. At the close of the experiment the respective percents were 85, 13 and 2. Thus it is apparent that there was practically no change in the number of pupils in each of the three sectional groups.

Table XXIV shows that the number of the duller pupils making more than normal progress was not very large in either group of schools. Slightly over one-half of the duller pupils of the experimental schools made regular progress in the slow sections, which resulted in their covering two semesters' work during the three semesters of the experiment. Slightly less than one-fourth of them made three semesters' progress by maintaining membership in the average sections. Ten percent managed to make more than normal progress, while 15 percent made less than two semesters. In the control schools 50 percent made normal progress, 45 percent less and 6 percent more. The average progress was three-tenths of a semester greater for the pupils of the control schools and the "progress quotient" fourteen points greater.

Analyzing the data presented above it is apparent that more of the duller pupils were failed in the control schools than in the experimental schools. Since, however, pupils were able to advance in the slow sections without failure while covering less than the normal amount of work the average progress was less in the experimental schools. As was true in the case of the brighter pupils more of the pupils from the experimental schools made extra progress.

The achievements of the duller pupils. The table just below gives the median achievement ages and quotients of the duller pupils of both groups of schools in February, 1921, and May, 1922.

	Experimental			Control		
	1921	1922	Gain	1921	1922	Gain
Achievement Age	8–10	10–6	20	9–3	10–9	18
Achievement Quotient	112	114	2	115	114	–1

These data show that the gain made by the duller pupils of the experimental schools was two months of achievement age greater

than that made by those of the control schools and that their gain in achievement quotient was three points greater. Therefore it can be said that in so far as achievement was concerned the experimental plan of procedure was slightly more efficient for the duller pupils than the traditional plan used in the control schools.

Summary. A special study of the brighter and duller pupils who were in school throughout the experiment yields the following results and conclusions:

1. The brighter pupils of the experimental schools had a "progress quotient" twenty-five points greater than did those of the control schools.
2. The relative gain of the brighter pupils of the experimental schools in median achievement quotient was three points.
3. The "progress quotient" of the duller pupils of the experimental schools was fourteen points less than that of the duller pupils of the control schools.
4. The duller pupils of the experimental schools made a relative gain of three points in their median achievement quotient.

A fair statement of the conclusions to be drawn would seem to be that the experimental plan of organization was considerably more efficient than the traditional plan in so far as it concerned the brighter pupils, but that in the case of the duller pupils it was somewhat less efficient.

CHAPTER VII

RESULTS AND CONCLUSIONS

A brief statement of the results of this experiment. The results actually obtained in this experiment may be listed as follows:

I. At the beginning of the experiment the placement of the pupils involved a relative loss in placement of .05 semester on the part of the experimental schools. This and the other amounts of progress are computed relative to the capacity of the pupils.

II. The main study, which included all the pupils of the two groups of schools, showed that:

1. The average progress was .06 semester larger in the experimental schools than in the control schools.
2. There was a relative gain for the experimental schools of about one point per semester in the achievement quotient as measured by the arithmetic and reading tests.
3. The achievement quotient derived from the Omnibus Test was thirteen points less for the experimental schools.

III. A special study of the pupils who remained in school throughout the experiment gave the following results:

1. The average progress for the experimental schools was .04 semester greater than that for the other group.
2. There was a relative gain for the experimental schools of one point in the achievement quotient.

IV. A special study of the brighter pupils revealed the following facts:

1. Those of the experimental schools progressed at a rate .25 semester greater than did those of the control schools.
2. The relative gain in the achievement quotient on the part of the pupils of the experimental schools was three points.

V. A special study of the duller pupils gave the following results:

1. Those of the experimental schools made, on the average, .14 semester less progress per semester than did those of the other group of schools.
2. There was a relative gain of three points in the median achievement quotient for the experimental schools.

It seems fair to summarize these results by saying that for pupils of all degrees of intelligence combined the experimental plan of organization was more efficient as regards progress and about the same as regards achievement, as compared with the traditional plan. The difference in progress was considerably more than enough to balance the relative loss caused by the placement of the pupils at the beginning of the experiment.

Conclusions to be drawn from these results and their application to school systems in general. The comparisons that were made between the schools taking part in this experiment and certain other city school systems seem to show that the results obtained in this experiment and the conclusions based thereon are fairly applicable to school systems in general. Assuming that this conclusion is warranted, the question remains as to whether or not the classification of pupils along lines similar to those followed in this experiment should be recommended to school administrators as a practical method of procedure. In considering this question it should be recognized that the public school superintendent or supervisor can ordinarily exercise a somewhat higher degree of supervision over the schools under his control than could the writer over the schools participating in this experiment. Therefore, it should be possible to secure somewhat more favorable conditions for carrying out the experimental plan of organization than were possible in this experiment.

In the second place, the question arises as to how large a gain in output, that is to say in progress and achievement, is required to justify a certain amount of additional investment. In this experiment the cost in both money and time was considerably larger per pupil than would be necessary in the usual public school situation. Ordinarily pupils would not need to be tested so often nor would it be necessary to use tests of achievement. Furthermore, there were many tabulations and computations made in this project that would not be necessary in the ordinary school situation. The cost

of group intelligence tests is only a few cents per pupil, in some cases being as low as one and one-half cents and in few more than ten cents. If the teachers scored the papers there would be no extra expense involved therein. Thus the cost of the tests and a rather small amount of clerk hire would be all the unusual outlay required to make use of group intelligence tests for purposes of placing pupils. Certain plans of doing this have involved a decrease in the average number of pupils per teacher or per room or some other element of additional investment. In this experiment there was no such expenditure, nor need there be in the usual situation. The desirability of reducing class size, whether in this or some other type of organization, is a separate problem. Therefore the total cost of the type of organization used in the experimental schools amounts to only a fraction of one percent of the total expenditure per pupil. As the gain in progress on the part of the experimental schools amounted to several percent of the total progress and as there was no loss in achievement, and, furthermore, as it is probable that under ordinary conditions the gain would be greater than it was in this project it would seem that an additional investment of a fraction of one percent would be entirely justifiable.

There remains, however, another point that must be considered in this connection. In Chapter I, output was defined as being composed of progress and achievement. There are undoubtedly other less tangible factors that constitute a part, and a rather important part, of the output of a school system. Such outcomes as industry, good citizenship, intellectual honesty, social development, etc., were either not measured in this experiment or measured so indirectly that no assumptions can be made concerning their presence and amount. This fact does not invalidate the conclusions reached, but merely signifies that these other outcomes of instruction must be considered in their interpretation. The fact that we cannot measure the total output should not bar us from measuring that which can be measured nor from proceeding according to what our measurements reveal until more complete measurements are possible.

It must also be remembered, as was stated in Chapter I, that there were really two problems involved in this experiment. It is possible that a portion or all of the results obtained in this experiment

might be secured in a somewhat similar experiment in which the pupils were classified according to teachers' judgments. Especially might this occur if the teachers participating were well-trained and experienced, and perhaps had given special study to the problem of classifying pupils according to their capacities. There were several reasons why the pupils in the control schools were not so classified, the chief one being that it was impracticable in the given situation. It may be suggested that since fast, average and slow sections were not formed in the control schools, they should not have been formed in the experimental schools. As was stated in Chapter I, the use of intelligence tests for the purpose of placing pupils implies that the pupils be placed according to their capacities and that it would not have been possible to arrange an experiment that would show the value of intelligence tests for the purpose mentioned unless such sections had been formed. Also the writer does not believe that the classification of the pupils of the experimental schools according to the teachers' judgments would have yielded as favorable results as did their classification according to the principles enumerated in Chapter I. This belief is based upon a study of the accounts of various experiments and of the teachers' estimates of capacity and the average school marks actually given in this experiment. These disagreed with the results of the intelligence tests in many cases and in most of these the latter appeared to furnish a more reliable means of predicting future progress and achievement than did the former.

In considering the conclusions reached from this study it should be borne in mind that the total time included was only three semesters. It is probable that if the experiment had continued for a longer time, say for eight or ten years, certain effects would have been noted that did not appear during the three semesters or effects that were present might have appeared in much more pronounced fashion. In general it seemed that as the experiment progressed from semester to semester the plan of organization being tried out gave better results. If the teachers had had several years' experience with such a plan the results might have been still more favorable. The plan was new to the teachers and hence they probably could not do their best work at first. On the other hand, it is possible that a division of the pupils into three groups

might tend to make the teachers feel less responsible for the achievements of the pupils, especially those of the duller ones. They might more or less unconsciously come to feel that the pupils placed in the slow sections could not be expected to do a very high quality of work and that therefore they were not worth much attention and effort. Such a result would, of course, be decidedly undesirable.

Considering the facts and possibilities mentioned above it is the opinion of the writer that the use of intelligence tests as the chief basis of classifying pupils increases the output of the school sufficiently to justify the additional expense involved. It is not, however, a panacea for all inefficient schools nor a method of organization that should be rushed into by every school administrator before he has made a careful study of its installation and operation.

APPENDIX A

A COMPARISON OF THE PUPILS ENTERING AND LEAVING SCHOOL DURING THE EXPERIMENT WITH THE TOTAL NUMBER OF PUPILS

Necessity for this comparison. Inasmuch as the shifting of membership within both groups of schools was so large, it seemed wise to take definite account of its effect upon the results and conclusions reached. In Chapters III and IV, where these results and conclusions are given, this effect has been considered. It was more or less probable that the number or mental capacities of the pupils eliminated from the experimental schools might be considerably influenced by the conditions of the experiment. For example, the recognition of the ability of the brighter pupils might tend to hold a larger percent of them in school and the placing of the duller pupils in slow sections might cause more of them to leave school than would normally be the case. This would, of course, materially raise the general mental level of the pupil material. On the other hand, it is possible that by placing many duller pupils, who would otherwise be failed, in the slow sections more of them would be held in school and that by allowing the brighter pupils to progress more rapidly they would be encouraged to leave school sooner than would otherwise be the case. Such results as these would lower the general mental level. Or perhaps some other combination of the four possible results just mentioned took place, so that more pupils of all degrees of ability were held in school, or more eliminated. Or again, other effects than those mentioned might have resulted. In regard to the new entrants, a priori reasoning would lead to the conclusion that they would have no effect upon the outcome of the experiment, since its operation would not in any way cause them to enter or not to enter school. However, it was thought best to make a study of them as well as one of the eliminees.

The effect of the pupils entering and leaving school during the experiment upon the total school population. It was found that during each of the three semesters of the experiment

the percent of pupils eliminated from the experimental schools was much greater than that from the control schools, the averages being about 12 and 7 percent, respectively. It might seem, therefore, that the experimental plan of organization resulted in increasing the amount of elimination. The writer does not believe, however, that this was the case. If it had been, the elimination rate for the pupils in the different sectional groups probably would have varied considerably. A study of this phase of the question shows that for each of the semesters the percents of all the pupils belonging to the fast, average and slow sections that were eliminated were practically the same. To word it differently, the percent of all pupils eliminated that had been in the fast sections was almost exactly the same as the percent of all pupils placed therein. A similar condition held for the other sections. Furthermore, the principals of the experimental schools stated that the elimination was no greater than was usual.

Table XXV shows the effects of the entrance and the elimination of pupils upon the total school population. It is to be read as follows, taking the first double column of the row of entries following "Med. Chron. Age" as an example: the elimination of pupils during the second semester of 1920-21 caused a decrease of one-tenth of a year more in the median chronological age of the pupils of the experimental schools than in that of the control schools. The entrance of new pupils during this time had no effect.

TABLE XXV. THE EFFECTS OF THE ENTRANCE AND ELIMINATION OF PUPILS UPON THE TOTAL PUPIL POPULATION

	Second Semester of 1920–21		Summer of 1921		First Semester of 1921-22		Second Semester of 1921-22	
	Elim.	NewE.	Elim.	NewE.	Elim.	NewE.	Elim.	NewE.
Med. Chron. Age	−.1		−.2		−.1	−.1	−.2	
Percent Accelerated	−1.		−1.	+1.			−1.	
Percent Retarded	+1.		+1.				+1.	
Aver. Retardation	−.03	+.03	−.04		−.02	+.02	−.04	
Aver. Progress		−.01		−.01		−.01		
Prog. Quotient		−1.						
Median M. A.			−.1			−.2	−.1	
Median I. Q.						−1.	+1.	
Median A. A.			−2.	−1.		−1.		
Median A. Q.				−1.				

In making use of the data in this table it must be borne in mind that all of the eliminated pupils were not included in the tabulations from which the data were derived. In a rather large number of cases the individual record cards of pupils who had left school were not returned to the writer along with the cards of those still in school. Practically all of these cases were in the control schools. In other cases the pupils were absent at the time of testing but did not actually withdraw from school until later, not returning to be tested in the meantime, so that another possible source of discrepancy was introduced. In view of these facts it was not certain that the effects listed in the table were all of the effects or were the true effects produced upon the pupil material by the pupils who left during the experiment. In the case of the new entrants there were no such opportunities for records to be lost unless the pupils concerned not only entered but left during the same semester, in which case they would not have been included in the tabulation.

The effect of the differences between the new entrants and eliminees and the total pupil population in so far as they relate to progress were considered in Chapter III. On the whole these effects were comparatively small. Those having to do with achievement were not used in Chapter IV or elsewhere. The reason for this was that all the tabulations in that chapter were made for the pupils who were present throughout the semester and hence did not need to be included for the pupils entering or leaving during the given semester. They are merely presented here as a matter of interest.

APPENDIX B

THE RELIABILITY AND CORRELATION OF THE TESTS USED IN THIS EXPERIMENT

In considering the results of such an experiment as the one described in the body of this report the question of the reliability of the tests used at once arises. The writer will not go into the matter in a detailed way but will merely present such coefficients of correlation and other measures of the reliability of the tests as were obtained and comment briefly thereon. No attempt was made to compute all the possible correlations between the tests used.

Constant and variable errors. Before proceeding to give the data referred to in the preceding paragraph, a brief discussion of the errors present in test scores seems appropriate. These errors may be classified as constant and variable.

Constant errors are those which are the same or approximately so for the group being tested. If, for example, the person giving the test allows less time than the directions call for a constant error is introduced, the effect of which is to lower the scores of all pupils taking the test. On the other hand, if too much time is allowed the scores are too large. Probably the most frequent constant errors are those due to what is often called "practice effect." If a duplicate form of a test is given the scores made thereon are ordinarily somewhat higher than those made at the first trial. Such constant errors were, of course, present in this experiment but as they were equally present for the two groups of schools it was not necessary to make any allowance for them.

Variable errors are those which differ for the different individuals taking the test. They are due to a number of causes. On any given day certain pupils are below par physically or mentally and therefore are likely to make a lower score than they would ordinarily. Such happenings as the breaking of a pencil point, the dropping of a test paper upon the floor or some occurrence distracting an individual's attention cause variable errors. All of these

mentioned so far result in lower scores. On the other hand, it may be that the particular form of a test used contains items which happen to be well known by a few members of the group taking the test. Such a condition results in an increased score. Scores may also be increased if a pupil turns the page and starts before the signal is given, if he does not know the correct answer but gets it by looking at someone else's paper, and by various other causes. It is usually impossible to determine the variable errors present in the scores of the individual pupils, although this can sometimes be done by a more or less detailed investigation. The effect of these errors is that the scores of many of the pupils are slightly too large or too small and those of a few are very much in error. On the other hand, the variable errors cause very little or no change in the average. In the long run they are as often positive as negative and therefore offset each other in the computation of averages.

The reliability of the Pressey Primer and the Illinois General Intelligence Scales. As the two scales named were the only ones used more than once in this experiment, they are the only ones for which the reliability can be calculated. The coefficients of correlation or of reliability,[1] the indices of reliability,[2] the probable errors of measurement,[3] and the percents these probable errors were of the respective medians were calculated.[4] Throughout the discussion of these measures of reliability it should be remembered that they were all computed from the use of tests at intervals of about six months and one year and therefore should not be expected to show as high a degree of reliability as if the time intervals had been shorter. In most studies of the reliability of tests the

---

[1]The coefficient of correlation between repetitions or duplicate forms of the same test is called the coefficient of reliability.

[2]The index of reliability is the square root of the coefficient of reliability. It measures the correlation between the score on one trial of a test and the true score. This true score is the average of the scores made upon an infinite number of trials of the test after these have been corrected for any constant errors.

[3]The term "probable error of measurement" bears the same relation to the index of reliability that the probable error of estimate bears to the coefficient of reliability. It is a measure of the variable error by which a pupil's score upon one trial of a test deviates from his true score. The formula is $.6745\sigma\sqrt{1-r}$. For σ the average of the standard deviations obtained from the scores made on each of two trials is used.

[4]The complete tables are to be found in the dissertation by the same title and author.

interval between the periods at which the tests were given has not exceeded a few days.

Table XXVI, Part A, shows that there was in general little difference in degree of reliability between the Pressey Primer and the Illinois General Intelligence Scale, that of the former being slightly higher. The average coefficient of reliability was in each case about four-tenths for the single half-grade groups and not far from seven-tenths for all grades combined. The average indices of reliability were somewhat greater than six-tenths and eight-tenths, respectively. The probable error of measurement averaged about nine points, or 15 percent of the median, in both cases. In the case of the Illinois Scale this amounts to almost one year of mental age, whereas in that of the Pressey it is somewhat less.

TABLE XXVI. DATA CONCERNING THE INTELLIGENCE TESTS USED IN THIS PROJECT

A. Reliability

	Coefficient of Reliability	Index of Reliability	Probable Error of Measurement	P.E. Meas. / Median
Pressey (Nov., 1920 and May, 1921)				
Grade Average..........	.46±.02	.67±.02	8	.16
Grades Combined........	.65±.01	.81±.01	9	.15
Illinois (Form 1 in Nov., 1920 and Form 2 in May, 1921)				
Grade Average..........	.38±.03	.62±.03	9	.15
Grades Combined........	.69±.01	.83±.01	9	.15
(Form 2 in May, 1921 and Form 1 in Dec., 1921)				
Grade Average..........	.46±.04	.67±.02	8	.14
Grades Combined........	.73±.01	.85±.01	9	.13
(Form 1 in Nov., 1920 and in Dec., 1921)				
Grade Average..........	.32±.04	.55±.03	9	.14
Grades Combined........	.62±.01	.79±.01	10	.15

B. Correlations Between the Different Tests Used.

	Grade Average	Grades Combined
Pressey and Dearborn (Used at Same Time)............	.59±.02	.78±.01
Illinois and National (" " " ")............	.63±.02	.81±.01
Pressey and Myers (" one semester apart)........	.29±.03	.39±.01
Kingsbury and Myers (" " " ")........	.47±.01	*........
Pressey and Illinois (" " " ")........	.20±.03	*........
Dearborn and Illinois (" " " ")........	.23±.03	*........
Pressey-Dearborn and Illinois (Used one semester apart).	.36±.03	*........
Dearborn and Myers (Used one year apart)...........	.38±.04	.52±.01
Pressey and Myers (" " " ")...........	.32±.04	.46±.01

*In these cases the correlations from only one grade are available.

In other words, the mental ages derived from a single application of the tests would be within that distance of the true mental ages in only about 50 percent of the cases.

Certain data as to the reliability of these two scales have been given by their authors. The administration of the Pressey scale to 365 first, second and third grade pupils gave an average coefficient of reliability of .92 between the first and second halves of the scale.[5] With two other groups of pupils numbering slightly over 100 each, coefficients of .89 and .92 were obtained.[6] The probable error of measurement was found to be between two and three points on the scale. These coefficients are naturally much higher and the probable errors much less than those obtained in this experiment because of the difference in the intervals between testing. The coefficients of reliability for Forms 1 and 2 of the Illinois scale are not quite as high as those between the two halves of the Pressey scale. Results based upon about 1000 children gave an average coefficient of .83 for grades III to VIII and one of .92 for the grades combined.[7] The probable error of measurement was between five and six points on the scale. These figures also show a considerably higher degree of reliability than do those obtained by testing at intervals of six months and one year. Inasmuch as the scale of the Illinois is finer than that of the Pressey, the probable errors are not far from the same when converted into mental ages.

The coefficients of reliability that are given for two or three other group intelligence tests run from about .75 up.[8] They tend

---

[5]Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades." Journal of Educational Psychology, 10, 297-308, September, 1919.

[6]Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades." Journal of Educational Research, 1, 285-94, April, 1920.

[7]Monroe, W. S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921. p. 47-49.

Monroe, W. S. and Buckingham, B. R. "The Illinois Examination I and II. Teacher's Handbook." Bloomington: Public School Publishing Company, 1920, p. 31.

[8]Colvin, S. S. "Educational Tests at Brown University." School and Society, 10, 27, July 5, 1919.

Colvin, S. S. "Some Recent Results Obtained from the Otis Group Intelligence Scale." Journal of Educational Research, 3, 1-12, January, 1921.

Otis, A. S. "An Absolute Point Scale for the Group Measurement of Intel-

to average about .80. Therefore, if these few are typical of similar tests in general, it would seem that the Pressey and Illinois scales are more reliable than are most group intelligence tests. From such a comparative standpoint coefficients of reliability around .90 and probable errors of measurement of two and five points may be said to be rather satisfactory. The differences between these figures and those obtained in this project may be largely, if not entirely, attributed to the difference in the time elapsing between the giving of the tests.

The correlations between the different group intelligence tests used in this experiment. Part B of Table XXVI presents the correlations obtained between the different tests used. It will be seen that the correlation between the Pressey scale and the Dearborn tests and that between the Illinois scale and the National tests are fairly high. An average correlation of about .60 when pupils are taken by half-grade groups and of about .80 for all grades combined is higher than is usually found between group intelligence tests.

The correlations between the results of the tests used at intervals of six months and one year are considerably lower. This would, of course, be expected as they take account not only of the differences between the tests but also of changes in the true mental abilities of the pupils during the period elapsing between the giving of the tests and of differences in the general conditions of testing at the two times. On the whole, these correlations do not compare unfavorably with similar correlations obtained elsewhere.

The writer collected data concerning the correlations found between different intelligence tests in some fifty cases. In practically all of these the different tests were given within a comparativly short time of each other, usually within the same week. The unweighted average of the coefficients of correlation was .62, which is only slightly higher than the average correlation by half-grade groups given in Part B of Table XXVI and much lower than that

---

ligence." Journal of Educational Psychology, 9, 333-47, and 237-61 May, 1918, and June, 1918.

Snarr, O. W. "Reliability of General Intelligence Tests in Classifying High School Pupils." Unpublished Thesis, University of Chicago, Chicago, June, 1919.

obtained for the grades combined. This is true although in a number of cases the coefficients were based upon several grades combined. Only about a dozen of the fifty are as high or higher than those of .78 and .81 which were obtained in this experiment when the grades were combined. In only one case was there a coefficient found higher than .90. Thus it may be said that the correlation between the Pressey scale and the Dearborn tests and that between the Illinois scale and the National tests were rather satisfactory as compared with similar correlations obtained in other experiments.

Although the coefficients given in Part B of Table XXVI were obtained from testing at intervals of one and two semesters, yet some of them compare favorably with a number of those given in the accounts of other experiments. When several half-grade groups were combined the coefficients averaged about .46.

The degree of reliability of single test scores was of concern in placing the individual pupils, but in measuring the results of the experiment this was not a matter of importance. The average used in most cases was the median, and for this the probable error is 1.25 (approx.) times the probable error of the distribution divided by the square root of the number of cases.[9] As the number of pupils included in this experiment was so large, the distribution would have had to be very scattering and the probable errors very large to cause the medians to be unreliable to any considerable degree. The distribution of the 3615 November, 1920, scores upon the Illinois scale, for example, had a probable error of 41 points, or 4.1 years of mental age. The probable error of the median was therefore about .85 point or one month.

[9]Yule, G. U. "An Introduction to the Theory of Statistics," London: Charles Griffin and Company, 1919, p. 338.

APPENDIX C

THE OMNIBUS TEST

As was mentioned in Chapter IV, a test called the Omnibus Test was devised by the writer to measure certain achievements of the pupils in the upper grades that were not covered by the reading and arithmetic tests used. This test was of the true-false type. It included seventy-five statements of which approximately half were correct and half incorrect. The following gives the first ten statements of the test:

1.	Russia produces a large amount of wheat..........	....
2.	The ancient Greeks were famous for their art...	....
3.	Charcoal is made from wood..........................	....
4.	4.6 is 100 times .46..	....
5.	A paragraph should be indented......................	....
6.	Italy raises a great deal of flax.......................	....
7.	The Roman Empire was not as powerful as Greece..	....
8.	Digestion begins in the mouth........................	....
9.	41/1000 = .41...	....
10.	A compound sentence has at least two independent clauses..	....

Every fifth statement had to do with the same subject, the five subjects included being geography, history, elementary science, arithmetic and grammar. The fifteen statements dealing with each subject were divided approximately equally between the six semesters of work covered and were in all cases based upon material mentioned in the outline of the Chicago course of study. The seventy-five statements were preceded by explicit directions and preliminary practise statements. The pupils were instructed to place a plus mark after those statements that were correct and a minus sign after those that were incorrect. A time limit of four minutes was placed upon the test.

BULLETIN NO. 13

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE STATUS OF THE SOCIAL SCIENCES IN THE HIGH SCHOOLS OF THE NORTH CENTRAL ASSOCIATION

by

WALTER S. MONROE
Director, Bureau of Educational Research

and

I. O. FOSTER
Instructor, University High School

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1922

TABLE OF CONTENTS

PREFACE

This report presents a summary of the information secured by means of a questionnaire addressed to the teachers of Social Science of the high schools of the North Central Association of Colleges and Secondary Schools. The authors have confined the report to a mere summary of the facts. Certain tabulations of the replies have not been included because it was felt that they were not sufficiently important to justify such detailed presentation. No effort has been made to interpret the facts or to summarize the comments made by a number of the persons who answered the questions.

Credit should be given Mr. Foster for originating the study and for preparing the questionnaire blank. He has also had a large part in the preparation of the manuscript for publication. The Bureau of Educational Research desires to acknowledge its indebtedness to all who cooperated by furnishing the information asked for.

WALTER S. MONROE, *Director.*

December 1, 1922.

THE STATUS OF THE SOCIAL SCIENCES IN THE HIGH SCHOOLS OF THE NORTH CENTRAL ASSOCIATION

CHAPTER I.

INTRODUCTION.

The purpose and method of this investigation. This study was undertaken for the purpose of collecting facts pertaining to the place occupied by the various social sciences in the curriculums of the high schools of the North Central Association of Colleges and Secondary Schools. In addition, certain questions relating to the general procedure of instruction were included. The specific questions asked are given on p. 8. The questionnaire blank was sent to 1273 secondary schools belonging to the North Central Association, and approximately one-half of these blanks were returned. Some were improperly filled out and had to be rejected. Others arrived too late to be included in the tabulations. The tables on the following pages are based upon the total of 475 replies.

The questionnaire blank. The questionnaire blank was printed as a four-page folder—eight and one-half by eleven. A portion of it, the general explanation, and the specific directions for answering each of the questions are reproduced here. Additional columns were added so that the following sixteen social science subjects were included in the blank:

Ancient History	Community Civics
Ancient and Medieval History	Civics
Medieval History	Vocational Civics
Medieval and Modern History	Economics
Modern History	Sociology
English History	Social Science
General History	Educational Guidance
American History	Vocational Guidance

Name..........Date.......... School..........Address..........	Ancient History		Ancient and Medieval History		Medieval History	
1. In what years and semesters are these subjects offered?						
2. If pupils from other years are admitted, state the years						
3. How many pupils were enrolled last semester?						
4. How many pupils are now enrolled?						
5. Is this a required subject?						
6. Is this subject generally liked by the students?						
7. Does the class study the entire period covered by the text?						
8. If not, what are the date limits?						
9. Is notebook work required?						
10. Is the stereopticon used?						
11. Do you plan to offer this course again?						
12. Check the courses in which current events are taught						
13. How many weeks are required to complete this course?						
14. How many minutes per week are spent in recitations?						
15. Is supervised study used often?						
16. Is the socialized recitation commonly used?						
17. How long have these subjects been offered?						
18. Check any courses that have proved unsuccessful						
19. Do you feel that more social science should be taught?						
20. Check the courses that you feel the pupils should be required to take?						
21. Give on line 1 for each subject: the author, the basic text, the date of its copyright, and the publisher. If supplementary texts are used give on line 2 the same information for the *most important*.	2	1	2	1	2	1

Explanation and direction for answering questions. On the inside of this folder a number of questions are asked concerning social science studies. Ancient History, Ancient and Medieval History, Medieval History, Medieval and Modern History, Modern History, English

History, Civics, Economics, and Sociology need no explanation. General History is a course from the beginning of history to the present time. American History may be only the history of the United States or it may be that of the entire western hemisphere. Community Civics is a course of civics that puts most of its emphasis on the local community. Vocational Civics acquaints the pupils with various callings or vocations. A detailed study of each may or may not be made. Social Science is a mixture of any of the social studies. It is to the social sciences what General Science is to the natural sciences. Educational Guidance here means a specific course which is intended to direct the pupil's studies along lines most profitable to him; it opens up and explains new fields to him. Vocational Guidance is the direction of the pupil toward those vocations which best suit his abilities and interests. The blanks are left for you to designate any other social study which your school offers.

Each question is explained below and specific directions are given for answering it. The numbers refer to the questions. It is suggested that all the blanks for question 1 should be filled in before you pass to question 2, etc. Each question should be answered under each subject that is offered by your school unless the following directions state otherwise.

1. Put the year in Roman numerals and the semester of the year in Arabic numerals. If Civics is offered only to Seniors in the second semester, under Civics and opposite the first question you will write IV-2, etc.

2. If pupils from other years or semesters are not permitted to take any subject, write under it and opposite question 2 the word "No." If all are admitted, write "All." If the course is open only to certain years or semesters other than expressed in 1, explain by giving the years and semesters, as in question 1.

3. Give the numbers as nearly exact as possible, as 348.

4. Give the numbers as nearly exact as possible, as 125.

5. Answer in the appropriate space, "Yes" or "No."

6. Answer in the appropriate space, "Yes" or "No."

7. Answer in the appropriate space, "Yes" or "No."

8. This applies only to history. If in Ancient History the work ends at 476, write "to 476" in the appropriate space. If the work in Medieval History extends from 800 to 1700, write in the appropriate space "800-1700." It is very important that this question be answered. English History, American History, and General History are assumed to cover the traditional period; if not, state your practice.

9. Answer "Yes" or "No."

10. Answer "Yes" or "No."

11. Answer "Yes" or "No."

12. Put a check (✓) or a cross (×) in the appropriate spaces.

13. Give the exact number, as 36.

14. Give the exact number, as 250.

15. Answer "Yes" or "No."

16. Answer "Yes" or "No."

17. Answer in numerals, as "2½" for two and one-half years. If the subject has been offered longer than three years, answer "3."

18. Either a check or a cross may be used. If the cause of failure is known, please state it on the last page of this report.

19. Answer "Yes" or "No."

20. Check, as in 12 and 18.

21. This question is very important. If in Ancient History the basic text is "Breasted, Ancient Times," write "Breasted, Ancient Times, 1916, Ginn and Company." If Webster's Ancient History is used for supplementary work, write on line 2 "Webster, Ancient History, 1913, D. C. Heath." Count as supplementary texts only books covering the whole subject, which are not library reference books.

Any additional information or remarks may be placed on the last page of this sheet. Any experience that will throw light on this problem is welcomed. Any syllabus or course of study that you may send us will be much appreciated.

General plan of tabulating the replies. As the replies were received from high schools varying widely in total enrollment, the following groups based upon total enrollment were formed:

Group	Enrollment	Number of schools
I	1–150	122
II	151–400	197
III	401–800	83
IV	801–above	73
		Total........475

When no entry was made in the column for a given subject it was assumed that this subject was not taught in the high school reporting. The number of high schools in each group not teaching a given subject is found in the last column of the first five tables. Although these numbers are not repeated in subsequent tables it is necessary to keep in mind the fact that the different subjects are not taught in all of the high schools. In a number of instances some of the questions were answered for a given subject but no entry made for other questions. In such cases it was assumed that the subject was taught in the high school but for some reason the one replying to the questionnaire blank either was unable to answer the particular questions or had overlooked them. In the first five tables column nine gives the number of schools not answering the first question for each subject. Similar columns are not given for all of the tables. In the case of those tables which consist of percents apparent discrepancies due to the failure of entries to total one hundred are caused by the fact that the question was not answered by certain schools reporting. In order to economize space the percent of such schools was omitted. However, in any case it can be obtained by adding the percents given and subtracting their sum from one hundred.

General plan of report. It is interesting and illuminating to consider the information collected from two points of view: first, by questions, and second, by subjects. For obvious reasons it was necessary to make the tabulations by questions. The more important tables are reproduced on the following pages. These will be described briefly in Chapter II and certain significant facts for each subject pointed out in Chapter III.

CHAPTER II.

SUMMARY OF QUESTIONS.

Status of social sciences in secondary schools. The answers to the first four questions[1] indicate the place occupied by the various social sciences in the curriculums of the secondary schools of the North Central Association. Tables I, II, III, IV, and V give the semester or semesters in which each of the subjects is taught. For example, thirty-eight schools report that Ancient History is taught during the first semester of the first year, thirty-five during the second semester, and so on. One significant thing about these tables is the variability of practice. Although there is a fair concentration of frequencies, most of the subjects are reported as being taught in practically all of the four years. The number of schools not teaching the subjects is also significant. With the exception of American History which is not taught in four schools and Civics which is not taught in fifty-seven schools, no social science subjects can lay claim to being universally taught. Only five subjects—Ancient History, American History, Civics, Economics and Modern History—are reported as being taught in more than 50 percent of the schools. In addition to the variation of practice shown by these tables, the answers to the second question, which are not summarized in this report, indicate considerable leniency in admitting students from other years.

The tabulations of the answers to questions 3 and 4 are not reproduced here. They show in the first group of schools a median enrollment of from twenty to thirty-six, except for the last five subjects. In the second group of schools the median enrollment varies from thirteen to fifty-six. In the third group the variation is from twenty-six to one hundred thirty-eight, and in the fourth, from forty-two to two hundred thirty-eight. In general, the enrollments for the second semester are slightly less. One significant thing is the rather extreme variation in the enrollment for a given subject within

[1] 1. In what years and semesters are these subjects offered? 2. If pupils from other years are admitted, state the years. 3. How many pupils were enrolled last semester? 4. How many pupils are now enrolled?

TABLE I. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS
(Group I. Enrollment, 1-150. 122 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	38	35	33	33	4	4	2	2	1	44
Ancient and Medieval History	6	7	18	19	3	3				87
Medieval History	1	2	7	4	5	2	1			108
Medieval and Modern History			25	24	25	25	3	3		67
Modern History	1	2	13	17	18	21	3	4		70
English History	1	1	1	1	7	7		1		109
General History	2	2	5	5	1	1				112
American History					33	14	74	37		3
Community Civics	17	11	1	3	3	2	1	6		84
Civics	1	2	1	2	7	25	26	64		13
Vocational Civics	4	4	3	2	1	1			1	115
Economics		1		3	17	8	29	18		56
Sociology				1	7	17	13	27		75
Social Science		1		1		2		1		120
Educational Guidance						2		3		119
Vocational Guidance	1		1	2	1	1	1		1	116

TABLE II. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS
(Group II. Enrollment 151-400. 197 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	46	44	59	50	5	4	1	1		93
Ancient and Medieval History	31	35	56	56	3	4	2	3		121
Medieval History	2	7	4	13	4	1				171
Medieval and Modern History		1	37	34	42	48	1	1		115
Modern History	1	2	47	48	75	72	9	11		83
English History	2	1	2		15	16	4	2		175
General History	3	2	11	12	5	5	3	3		180
American History					39	24	166	105		0
Community Civics	41	38	9	10	2	5	1	6		122
Civics	4	3	6	3	28	40	51	116		26
Vocational Civics	4	6	1	2			1	1		182
Economics	1	1	4	4	31	41	50	60		68
Sociology	1	1	2	2	14	22	26	37		130
Social Science	1	2	1	1	1		5	6		185
Educational Guidance									3	194
Vocational Guidance	2	5	2	1			1			187

TABLE III. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group III. Enrollment, 401-800. 83 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	22	17	21	17	1	1	1			42
Ancient and Medieval History	18	19	28	29	3	3				45
Medieval History	1	3	3	8	5	5	1	1		67
Medieval and Modern History		1	19	14	12	12	2	1		54
Modern History			21	21	32	31	2	2		32
English History	1	1	2	2	10	9	2			68
General History	1	1	7	7	1	1		1	1	77
American History			2	1	22	20	64	39		1
Community Civics	18	17	1	1			1	2		55
Civics			4	6	16	15	22	52		10
Vocational Civics	2	3		1					2	75
Economics			2	2	11	14	27	31		21
Sociology			2	1	2	8	10	12		58
Social Science	1	1					1	2	1	77
Educational Guidance	1						1		1	80
Vocational Guidance		2	2						1	78

TABLE IV. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Group IV. Enrollment 801 and over. 73 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	16	12	15	18	2	2	2	2	1	36
Ancient and Medieval History	16	16	16	16	3	3				39
Medieval History	1	7	5	3	9	3				51
Medieval and Modern History		1	7	8	9	8	1	1		52
Modern History			21	22	23	27	7	4		19
English History			3	4	14	17	2	2		52
General History			2	2	1	1				67
American History			1	1	13	13	54	49		0
Community Civics	17	11	4	4			3	3	1	40
Civics			4	4	11	9	35	45	1	8
Vocational Civics	4	2	3	1	1				1	65
Economics				1	11	12	30	37		14
Sociology		1			3	4	8	11	1	56
Social Science			5	3			4	3		62
Educational Guidance										73
Vocational Guidance	2	1	3	3			1	1		67

TABLE V. SHOWING YEARS AND SEMESTERS IN WHICH VARIOUS SOCIAL SCIENCES ARE TAUGHT IN HIGH SCHOOLS

(Total. 475 High Schools Reporting)

Subject	Year and Semester								No Answer	Not Taught
	I		II		III		IV			
	1	2	1	2	1	2	1	2		
Ancient History	122	108	128	118	12	11	6	5	2	215
Ancient and Medieval History	71	77	118	120	12	13	2	3		292
Medieval History	5	19	19	28	23	11	2	1		397
Medieval and Modern History		3	88	80	88	93	7	6		288
Modern History	2	4	102	108	148	151	21	21		204
English History	4	3	8	7	46	49	8	5		404
General History	6	5	25	26	8	8	3	4	1	436
American History			3	2	107	71	358	230		4
Community Civics	93	77	15	18	5	7	6	17	1	301
Civics	5	5	15	15	62	89	134	277	1	57
Vocational Civics	14	15	7	6	2	1	1	1	4	437
Economics	1	2	6	10	70	75	136	146		159
Sociology	1	2	4	4	26	51	57	87	1	319
Social Science	2	4	6	5	1	2	10	12	1	444
Educational Guidance	1					2	1	3	4	466
Vocational Guidance	5	8	8	6	1	1	3	1	2	448

a group of schools. Differences in the total enrollment must, of course, be kept in mind, but even so some of the variations seem of especial interest. For example, in Group I two schools reported an enrollment in Ancient History of between one hundred and one hundred twenty-five. Since no schools in this group have a total enrollment greater than one hundred fifty, all but a very few students in these two schools were studying Ancient History. Similar conditions for Ancient History as well as for other subjects are found in the other groups. This suggests that in many cases the course of study is in a state of change, or that distinctly different policies with reference to educational guidance are being followed in different schools.

The scope and content of the various social sciences. Questions[2] 7, 8, 12, 19 and 21 furnish information relative to the content

[2] 7. Does the class study the entire period covered by the text? 8. If not, what are the date limits? 12. Check the courses in which current events are taught? 19. Do you feel that more social science should be taught? 21. Give on line 1 for each subject: the author, the basic text, the date of its copyright, and the publisher.

TABLE VI. PERCENT OF HIGH SCHOOLS IN WHICH THE CLASSES STUDY THE ENTIRE PERIOD COVERED BY THE TEXT

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.	Text Fol.	Text not Fol.
Ancient History	94	5	91	5	93	2	84	8	91	5
Ancient and Medieval	91	6	97	3	97	3	91	6	95	4
Medieval History	79	21	77	12	88	12	68	23	77	17
Medieval and Modern	96	4	93	3	90	3	86	5	93	4
Modern History	92	4	98	1	98	2	96	4	97	2
English History	84	8	86	5	86	7	95		89	4
General History	90		94	6	83		83		89	3
American History	96	3	90	7	92	4	93	6	93	5

TABLE VII. PERCENT OF HIGH SCHOOLS IN WHICH CURRENT EVENTS ARE TAUGHT
(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Percent of High Schools				
	Group I	Group II	Group III	Group IV	Total
Ancient History	30	45	54	49	43
Ancient and Medieval	40	47	55	47	48
Medieval History	29	38	50	55	44
Medieval and Modern	58	69	69	57	64
Modern History	60	78	75	78	74
English History	54	45	53	62	54
General History	40	47	67	17	44
American History	86	88	87	82	86
Community Civics	55	73	79	76	71
Civics	83	80	82	69	79
Vocational Civics	43	33	25	100	47
Economics	51	58	56	63	57
Sociology	53	52	44	47	51
Social Science	50	50	50	36	45
Educational Guidance	33		33		22
Vocational Guidance	17	10	20		11

of the various courses. The replies to question 7 given in Table VI indicate, except in the case of Medieval History, a very close adherence to the text books in regard to date limits.[8] The particular date limits for the various subjects will be noted in the next chapter devoted to the separate subjects. Table VII shows that current events are reported as being taught in a considerable percent of the classes in all subjects. As one might expect the percents are greatest for American History and for the various courses in Civics. A variety of texts is used, although there is considerable agreement upon a limited

[8]In interpreting the facts in Table VI it is necessary to bear in mind that the percents have been computed on different bases in the case of the different subjects. For example, by reference to Table I we learn that Ancient History is taught in seventy-eight schools in the first group. Ancient and Medieval History is taught in thirty-five, Medieval in fourteen, Medieval and Modern History in fifty-five, etc. When it is stated in Table VI that 94 percent of the schools in Group I studied the entire period covered by the textbook in Ancient History, this simply means that this occurs in 94 percent of the seventy-eight schools in which Ancient History is taught. The 91 percent for Ancient and Medieval History is computed on thirty-five as a basis and the 79 percent for Medieval History is computed on fourteen as a basis.

The percents given in the remaining tables have been computed in the same way. Hence, it will be necessary to bear in mind this precaution in interpreting them.

number. This subject also will be discussed further in Chapter III. The answers to question 19 and the general remarks made upon the questionnaire seem to indicate (1) that in the teaching of history less time should be spent on the political side and more on the economic, social and industrial phases, and (2) that in the teaching of the other social sciences the abstract and theoretical topics should be made more concrete and practical.

Time devoted to social science courses. In a few schools which appear to be organized on a three-term basis as few as twelve weeks are given to certain of the social science studies. At the other extreme we find forty weeks being devoted to certain of the subjects. The number of minutes per week spent in recitation varies from 120 to 435. Usually a balance between the number of weeks devoted to the subject and the number of minutes per week is affected by requiring a class that spends relatively few minutes per week in recitation to carry a subject for a compensating longer period. For this reason the variations in time devoted to the different subjects is not as great as it appears from the above figures. The typical social science course extends through two semesters of eighteen weeks in length, and requires 225 minutes per week for recitation. Larger schools have a tendency to be in session longer than the smaller schools; therefore, they devote in most cases more time to the social science work.

The social sciences as required subjects. The replies to questions[4] 5 and 20 acquaint us with the status of the various social sciences as required subjects. Table VIII, which gives a summary of the replies to question 5, indicates a considerable difference of practice except in the case of American History and Civics. For the other subjects the predominant practice is to make them elective. The directions for replying to this question required an answer of either "yes" or "no". It is not unlikely that in a number of schools a subject is required for students pursuing certain courses but elective for others. As no opportunity was given to indicate this fact the number of affirmative replies probably was decreased. On the other hand, for the same reason an affirmative reply in some cases may have been slightly inaccurate. This character of the response required in this

[4]5. Is this a required subject? 20. Check the courses that you feel the pupils should be required to take.

TABLE VIII. PERCENT OF HIGH SCHOOLS IN WHICH VARIOUS SOCIAL SCIENCES ARE REQUIRED
(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.
Ancient History	49	47	29	62	42	51	30	59	37	56
Ancient and Medieval	51	40	24	67	23	66	20	68	28	62
Medieval History	29	71	38	54	37	44	27	64	33	58
Medieval and Modern	33	58	24	67	24	66	29	52	27	63
Modern History	34	60	27	67	35	59	43	46	33	60
English History	23	62	9	91	13	67		90	10	80
General History	50	30	29	59	67	33	50	17	44	41
American History	85	11	85	11	79	17	89	6	85	11
Community Civics	50	37	46	43	46	36	46	33	47	39
Civics	79	14	72	22	71	26	66	19	73	20
Vocational Civics	30	70	33	53	38	50	75	25	42	50
Economics	45	47	16	74	24	70	13	68	23	67
Sociology	49	49	18	71	16	80	6	70	26	66
Social Science		50	17	83		67	27	46	16	65
Educational Guidance			34	66	33	67			22	33
Vocational Guidance	17	50	10	70	60	20	67	33	33	48

TABLE IX. PERCENT OF HIGH SCHOOLS WHICH FAVOR MAKING THE VARIOUS SOCIAL SCIENCE STUDIES REQUIRED
(Based Upon Only Those Schools in which the Subject is Taught)

Subject	Percent of High Schools				
	Group I	Group II	Group III	Group IV	Total
Ancient History	35	31	24	14	28
Ancient and Medieval	43	17	24	21	24
Medieval History	14	38	44	14	28
Medieval and Modern	44	34	17	14	32
Modern History	50	51	59	44	51
English History	31	36	7	14	23
General History	60	65	67	50	62
American History	88	86	87	86	87
Community Civics	74	67	61	79	70
Civics	81	82	77	69	79
Vocational Civics	14	40	25	38	32
Economics	59	61	42	37	53
Sociology	62	45	28	53	48
Social Science		42	33	55	42
Educational Guidance	67	33	67		56
Vocational Guidance	50	40	60	50	48

question should be borne in mind in interpreting Table VIII. In Table IX which summarized the replies to question 20 we find that in the opinion of the persons reporting some slight changes should be made with respect to requiring the different subjects. However, in general these changes were not marked. It is probably significant that in both Tables VIII and IX a decrease in the requirement of certain subjects is indicated as the size of the high school increased. In other subjects there is a little change and for some there is even an increase.

Permanency and success of the various social sciences. Questions[5] 6, 11, 17 and 18 give some indication of the permanency and the success of the various social sciences from the standpoint of pupils' interest. Table X, which summarizes the replies to question 6, indicates that, in the judgment of the persons reporting, these subjects are generally liked by a majority of students. Three-fourths or more of the schools replied affirmatively for most subjects. In a con-

[5]6. Is this subject generally liked by the students? 11. Do you plan to offer this course again? 17. How long have these subjects been offered? 18. Check any courses that have proved unsuccessful.

TABLE X. PERCENT OF HIGH SCHOOLS IN WHICH THE VARIOUS SOCIAL SCIENCES ARE GENERALLY LIKED BY STUDENTS

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked	Liked	Not Liked
Ancient History	63	24	75	16	71	19	68	8	70	18
Ancient and Medieval	68	26	84	13	76	13	65	9	76	14
Medieval History	79	21	73	19	69	12	64	14	70	17
Medieval and Modern	67	24	85	9	72	14	71	5	76	14
Modern History	83	15	77	11	84	12	66	4	77	11
English History	92		77	5	73	7	71	5	78	4
General History	70		82	6	83	17	50	17	74	8
American History	90	3	92	3	93	2	76	3	89	3
Community Civics	87		90	1	89		70	3	85	1
Civics	82	4	83	9	82	4	66	6	80	7
Vocational Civics	100		67		88		100		84	
Economics	83	8	84	8	82	8	71	5	81	7
Sociology	92	2	84	1	92		88	6	88	2
Social Science	100		75	8	67	16	73		74	7
Educational Guidance	100					33			33	11
Vocational Guidance	67		60		80		67		67	

TABLE XI. PERCENT OF HIGH SCHOOLS IN WHICH NOTEBOOK WORK IS REQUIRED FOR THE VARIOUS SOCIAL SCIENCES

(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.	Req.	Not Req.
Ancient History	81	17	74	17	76	19	81	16	78	17
Ancient and Medieval	77	20	80	19	74	26	68	20	77	19
Medieval History	86	7	65	23	81	19	77	14	75	17
Medieval and Modern	76	20	77	16	79	14	61	29	75	18
Modern History	75	23	75	20	86	14	74	19	76	19
English History	62	31	77	14	67	20	81	10	73	17
General History	20	40	71	23	67	33	67		56	26
American History	75	21	75	20	75	23	80	19	75	21
Community Civics	42	40	48	41	46	50	67	24	50	39
Civics	60	32	56	36	59	38	60	31	58	35
Vocational Civics	57	43	40	33	38	62	75	25	50	39
Economics	46	41	39	49	44	50	49	42	43	47
Sociology	55	41	40	39	36	56	35	65	44	45
Social Science	50	50	58	42	50	33	55	27	55	35
Educational Guidance	33	67	33	67		33			22	56
Vocational Guidance	33	17	50	40	40	40	100		56	26

TABLE XII. PERCENT OF HIGH SCHOOLS IN WHICH THE STEREOPTICON IS USED
(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	26	68	22	64	34	63	89		35	56
Ancient and Medieval	23	74	24	70	11	87	35	62	23	73
Medieval History	22	71	23	62	44	56	50	41	35	56
Medieval and Modern	16	69	18	70	21	69	52	38	22	66
Modern History	29	67	19	70	22	78	35	57	25	68
English History	15	69	18	68	27	60	38	43	25	59
General History	10	60	47	53		67	33	17	28	51
American History	23	64	18	71	24	70	41	48	24	65
Community Civics	16	66	19	71	36	50	37	45	24	62
Civics	14	72	12	73	12	77	27	62	14	72
Vocational Civics	29	71	20	53	38	62	25	75	26	63
Economics	11	77	11	74	15	79	19	69	13	75
Sociology	21	70	15	66	12	80	12	76	16	71
Social Science		50	34	58	17	50	27	64	26	58
Educational Guidance	33	33				33			11	22
Vocational Guidance	17	33	10	80	20	60	16	67	15	63

TABLE XIII. PERCENT OF HIGH SCHOOLS IN WHICH THE SOCIALIZED RECITATION IS COMMONLY USED
(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	46	42	37	47	46	51	24	73	38	50
Ancient and Medieval	43	49	40	54	50	42	41	53	43	50
Medieval History	50	50	35	46	44	44	23	73	36	54
Medieval and Modern	40	55	88	12	31	59	24	67	58	32
Modern History	50	46	43	49	50	40	41	54	45	48
English History	23	46	36	50	33	40	29	57	31	49
General History	30	50	23	65	17	50	33	50	25	57
American History	52	40	52	40	51	43	45	48	51	42
Community Civics	50	34	60	33	39	50	42	49	51	39
Civics	56	35	54	38	47	47	40	51	51	41
Vocational Civics	86	14	33	40	50	25	38	62	47	37
Economics	47	42	55	38	37	50	55	38	47	44
Sociology	58	36	54	43	60	36	65	35	57	39
Social Science	50	50	58	42	33	33	36	45	45	42
Educational Guidance	67	33		33		33			22	33
Vocational Guidance	17	33	50	40	40	40	33	67	37	44

TABLE XIV. PERCENT OF HIGH SCHOOLS IN WHICH SUPERVISED STUDY IS FREQUENTLY USED
(Based Upon Only Those Schools in Which the Subject is Taught)

Subject	Group I		Group II		Group III		Group IV		Total	
	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used	Used	Not Used
Ancient History	45	45	36	61	46	51	33	62	40	55
Ancient and Medieval	46	51	47	50	47	42	29	62	44	51
Medieval History	43	57	27	58	44	44	55	45	41	51
Medieval and Modern	35	63	29	67	41	59	29	62	33	64
Modern History	44	54	37	58	43	55	35	61	39	57
English History	31	46	23	64	47	33	19	81	28	59
General History	50	50	29	65	50	50	16	67	36	59
American History	34	57	34	64	40	56	29	66	34	61
Community Civics	37	50	36	56	43	43	42	46	38	51
Civics	32	61	27	68	37	60	29	65	31	64
Vocational Civics	29	71	53	33	13	50	25	62	34	50
Economics	30	59	28	67	37	57	27	71	30	64
Sociology	41	55	27	67	48	40	35	65	35	59
Social Science	50	50	25	75	17	67	45	55	32	65
Educational Guidance	33	67				33			11	34
Vocational Guidance	17	33	10	90	20	60	50	33	22	59

siderable number of instances this question was not answered. The failure to do so probably means that the person replying to the questionnaire was very doubtful concerning the attitude of the students toward the subject.

Methods of instruction. Questions[6] 9, 10, 15 and 16 give certain information concerning the general procedure of instruction. The replies to these questions are summarized in Tables XI, XII, XIII and XIV. It is rather surprising to the writers that notebook work is reported as being required in such a large percent of the schools. As no attempt was made to define "notebook work" in the questionnaire blank the amount required in some of the schools probably is not large. The use of the stereopticon, as one might expect, is more frequently reported in the larger schools than in the smaller ones. Tables XIII and XIV give the summary for the socialized recitation and supervised study respectively. These phases of instructional procedure were not defined and the persons responding to the questionnaire no doubt differed in their interpretations. This naturally limits the significance of the replies. It is probably significant that the use of the socialized recitation does not appear to be restricted to any particular subjects, although, with the exception of Sociology, the highest relative frequencies are for American History and Civics. The percent of schools reporting the frequent use of supervised study is approximately one-third of the total number. It appears that supervised study occurs more often in smaller schools than in the larger ones.

[6]9. Is notebook work required? 10. Is the stereopticon used? 15. Is supervised study used often? 16. Is the socialized recitation commonly used?

CHAPTER III.

SUMMARY BY SUBJECTS

Ancient History. Ancient history is reported as a separate course in more than 50 percent of the schools in each group. In addition, a considerable number of schools combine it with medieval history. It is slightly more frequently taught during the second year than the first, and in a few schools it is given during the third and fourth years. The enrollment compares favorably with that in the other social science subjects although it is usually exceeded by American History. The textbooks and the supplementary texts mentioned most frequently are given below:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Robinson and Breasted	History of Europe	49
West	Ancient World	39
Webster	Ancient History	32
Breasted	Ancient Times	30
Myers	Ancient History	24
Webster	Early European History	23
Betten and Kaufman	Ancient World	9
Ashley	Early European Civilization	9
West	Early Progress	8
	SUPPLEMENTARY TEXTS	
Davis	Readings	26
West	Ancient World	6
Myers	Ancient History	6
Breasted	Ancient Times	6
Webster	Ancient History	5
	Total number of schools offering Ancient History	260

The answers to question 7 indicated that in approximately 90 percent of the schools offering Ancient History the textbook was followed with respect to date limits. When the date limits of the text are not adopted, 500 A. D. and 800 A. D. are most frequently mentioned. Only a few extend Ancient History to about 1000 A. D. Current events are reported as being taught in Ancient History courses in slightly more than two-fifths of the schools. More attention ap-

pears to be given to this topic in the larger schools than in the smaller ones. In more than one-third of the schools in which Ancient History is taught it is a required subject. This requirement is a little more frequent in the smaller schools, and the percent of replies indicating a belief that Ancient History should be required is somewhat less. Thus, it appears that if the persons replying to the questionnaire were able to assign Ancient History to the place which in their opinion it should occupy it would be required in fewer schools than at present.

Ancient History is reported as being generally liked by students in 70 percent of the schools offering it, with 12 percent not reporting. This is the smallest percent reported for any social science except "Educational Guidance" and "Vocational Guidance" which are taught in too few schools for the data to be significant. Apparently very few schools are planning to discontinue Ancient History as a separate course. Notebook work is required in about three-fourths of the schools. Use of the stereopticon is more frequent in Ancient History classes than in any other social science except Medieval History in which the same percent reports its use. The socialized recitation is used somewhat less frequently in Ancient History than in most of the other social science studies but in respect to supervised study it stands near the head of the list.

Ancient and Medieval History. This new combination of history is designed to take the place of the full year of Ancient History and one-half year of Medieval History in order that more time may be devoted to Modern History. About one-third of the schools reporting have adopted this combination in preference to the traditional one of Ancient History in the first or second year of the high school, followed immediately by Medieval and Modern History in the second or third year. The normal place for this combination subject is the second year, although it is occasionally reported for the third and fourth years. In schools having the same number of pupils the enrollment in this subject is reported as slightly larger than that in Ancient History.

Since the textbook defines the limits of the course in nearly every case the reader's attention is called to the list of textbooks and of supplementary texts that are most frequently used.

PRINCIPAL TEXTS

Author	Title	Frequency
Robinson and Breasted	History of Europe	72
Webster	Early European History	28
Robinson and Breasted	Outlines of European History	24
Breasted	Ancient Times	15
Robinson	Medieval and Modern Times	13
Ashley	Early European Civilization	8
	SUPPLEMENTARY TEXTS	
Breasted	Ancient Times	6
Davis	Readings	5
Total number of schools offering Ancient and Medieval History		183

The final date limit of this course appears to be somewhere between 1476 and 1800, the year 1500 being given in most cases. In this combined course current events are taught more frequently than in the year course in Ancient History. More of the smaller schools than the larger require this subject. It has been offered in 124 schools for three years or longer, and in 28 schools for less than three years. Very few administrators plan to discontinue the course.

Medieval History. A critical examination of the data on this subject shows that Medieval History has probably been recorded separately in some cases when it should have been reported with either Ancient History or Modern History. However, no account has been taken of this error in the following discussion. Medieval History is taught most often during the second semester of the second year or the first semester of the third year. Usually pupils enrolled in the year above that for which the course is offered are admitted to it. In about 80 percent of the schools the textbook sets the limits of the course. However, it seems that as a rule the chronological period covered extends from the Germanic Invasions to about 1700. The textbooks used in this course are as follows:

PRINCIPAL TEXTS

Author	Title	Frequency
Robinson and Breasted	History of Europe	12
Robinson and Beard	Outlines of European History	12
Robinson	Medieval and Modern Times	11
Webster	Early European History	10
	SUPPLEMENTARY TEXTS	
Robinson	Readings	5
West	Modern Progress	3
Hazen	(Not Mentioned)	2
Total number of schools offering Medieval History		78

Medieval History is required in about one-third of the schools in which it is taught. Usually these are the small schools which necessarily have a minimum of elective subjects. Hence, it is not surprising to find that slightly less than one-third of the persons replying to the questionnaire feel that this subject should be required. Except in the small schools it is not liked by the students as well as Ancient History. Only one school reported this course as unsuccessful and hence plans to discontinue it.

Medieval and Modern History. This subject is most frequently found in either the second or the third year of the high school. Only about one-third of the schools limit it to those pupils who have standing equivalent to the year in which it is offered. In most cases the smaller schools follow the textbook more strictly than the larger ones. About 27 percent of the schools replying to the questionnaire begin the work with the Decline of the Roman Empire, about 50 percent with the Germanic Invasions, and about 23 percent with the year 1000 A. D. More stress appears to be given current events than in any of the preceding subjects. The texts most frequently used are given below:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
West	Modern Progress	27
West	Modern World	19
Robinson and Beard	Outlines of European History	18
Harding	Medieval and Modern History	16
Myers	Medieval and Modern History	11
	SUPPLEMENTARY TEXTS	
Hazen	Europe Since 1815	12
Robinson	Readings	7
West	Modern Progress	5
Robinson and Beard	Outlines of European History	5
Total number of schools offering Medieval and Modern History		187

The replies indicate that the majority of the persons who filled in the questionnaire feel that this subject should be more universally required. About three-fourths of the schools report the subject as generally liked, about 7 percent state that it is unsuccessful, and about 4 percent expect to discontinue the course. Approximately 75 percent of the schools require notebook work, and the stereopticon is used in about 22 percent, the highest frequency being in the larger

institutions. Supervised study and the socialized recitation are found in approximately one-third and three-fifths of these schools respectively.

Modern History. This subject is usually offered in the second or third year, the frequency being highest in the third year. Occasionally it is taught in the first or the fourth year. In one school practically all of the students are reported as being enrolled in the course. The textbook is more nearly followed in this subject than in any other of the social sciences. Those used most frequently are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Robinson and Beard	Outlines of European History	64
Robinson and Beard	Our Own Times	62
Hazen	Modern European History	28
West	Modern Progress	20
Robinson	Medieval and Modern Times	15
Ashley	Modern European Civilization	12
West	Modern World	12
Webster	Modern European History	10
SUPPLEMENTARY TEXTS		
Hazen	Modern European History	9
Robinson and Beard	Our Own Times	8
Robinson	Readings	6
Total number of schools offering Modern History		271

In the smaller schools this course usually begins with the fifteenth century, but in the larger schools in most instances a later date is chosen. Current events are taught more frequently in connection with this course than with any other history course, except American History. It is interesting to note that about one-third of the schools require this subject for graduation, and that over one-half of the persons who answered the questionnaire think that it should be required. It has been offered for three years or longer in 190 schools, and for less than three years in 35 schools. It is more generally liked than Medieval and Modern History. Nine schools report it as unsatisfactory and six of these will not offer it again. Notebook work is required in about three-fourths of the schools. The data show that supervised study is in vogue in about 40 percent of these schools, and that the socialized recitation seems to be somewhat more widely used.

English History. This survival of the recommendation of the Committee of Seven is still found in over one-fourth of the high schools of the North Central Association. The usual year for this subject is the third, although sometimes it is offered in the second and occasionally in the first. When this subject occupies its normal position seniors usually are admitted and sophomores occasionally. No school reported that its pupils studied less than the entire period covered by the text. The texts most frequently used are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Cheyney	Short History of England	40
Andrews	Short History of England	5
Cheyney	Readings	5
Larson	Short History of England	3
Walker	Essentials of English History	3
	SUPPLEMENTARY TEXTS	
Cheyney	Short History of England	2
Larson	Short History of England	2
Cheyney	Readings	2
Gardiner	History of England	2
	Total number of schools offering English History	71

Current events are taught in slightly more than 50 percent of the classes in English History. The smaller schools which offer this course report that it is liked best of all the social subjects. It is comparatively well liked in the larger schools. In three schools it was reported as unsuccessful.

General History. This subject is offered in less than 10 percent of the high schools of the North Central Association. It is found usually in the second year, although it was reported for all years. It is generally taught for the entire year and pupils from the year just above or below that in which the subject is scheduled are usually admitted. Only one school reported a failure to follow the work as outlined in the textbook. The textbooks which have the highest frequencies are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Elson	Modern Times and the Living Past	10
Robinson-Breasted-Smith	General History of Europe	5
West	Modern Progress	5
Myers	General History	4
	SUPPLEMENTARY TEXTS	
Elson	Modern Times and the Living Past	3
West	Modern Progress	2
	Total number of schools offering General History	39

This course includes current events in more than 50 percent of the schools teaching it. It is interesting to note that in 44 percent of these schools the subject is required, and that the administrators in 62 percent of the schools think that it should be. This subject has been introduced into nineteen schools within the past three years. It is reported as being fairly well liked, and no schools pronounced it unsuccessful.

American History. American History is scheduled most frequently in the fourth year, although it is often found in the third. It is the only social science that is not offered by any school during the first year. Usually, pupils in the year above or below that in which the course is scheduled are admitted. When this course is a semester subject it is offered in the smaller schools most frequently during the first semester, but in the larger during the second semester. About 93 percent of the schools cover the entire period with which the text deals. It is probable also that they supplement the text with more recent history. The textbooks most frequently used are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Muzzey	American History	195
Fite	History of United States	47
Forman	Advanced American History	33
West	American History	30
McLaughlin	History of the American Nation	30
Hart	American History	22
	SUPPLEMENTARY TEXTS	
Bassett	Short History of the United States	18
Muzzey	American History	14
Hart	New American History	7
	Total number of schools offering American History	471

Only one school did not report current events in connection with this course. American History is required in more than 80 percent of these schools. This is approximately the same as the percent of returns indicating that it should be a required subject. It is reported as being well liked in about 97 percent of the schools where it is taught, and as unsuccessful in slightly less than 3 percent of these schools. No school, however, expects to discontinue the course next year. Of these schools notebook work is required in 75 percent; the

stereopticon is used in 24 percent; supervised study is found in 34 percent; and the socialized recitation finds a place in 51 percent. One significant and illuminating opinion, reported frequently, is that American History can not be done satisfactorily in one semester.

Community Civics. Community Civics is found most frequently in the first year and usually in the first semester. However, it is reported for every year of the high school. Few schools admit pupils from years other than that for which the course is scheduled. About 75 percent of the schools offering this course do the work as outlined by the texts. The texts in most frequent use are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Hughes	Community Civics	96
Dunn	The Community and the Citizen	17
Hill	Community Life and Civic Problems	11
Hughes	Economic Civics	6
	SUPPLEMENTARY TEXTS	
Hughes	Community Civics	5
Hill	Community Life and Civic Problems	5
	Total number of schools offering Community Civics	174

Only two schools do not teach current events in connection with Community Civics. In at least one case this is due to the fact that current events are taught in connection with other courses, and this practice was resorted to in order to avoid duplication. Community Civics has been in the curriculums of eighty-two schools for three years or longer. During the past three years, however, it has been introduced into the curriculums of seventy-three schools. Only 2 percent of the schools report that the subject is not generally liked. Eight schools do not consider it successful and one plans its discontinuance. Notebook work is required in connection with this course in about 50 percent of these schools. The socialized recitation is used by an equal number. Supervised study is not found as frequently, and the use of the stereopticon is limited to about 25 percent of the schools offering the subject.

Civics. This course remains predominantly a fourth-year subject, although it is taught in a large number of schools during the third year. The data indicate that about 85 percent of all the schools

offering courses in Civics follow the text rather literally. The textbooks most frequently used are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Magruder	American Government	129
Guitteau	Government and Politics, Etc.	50
Ashley	New Civics	46
Woodburn and Moran	Citizen and Republic	39
Reed	Form and Functions of American Government	23
Garner	Government in the United States	18
Forman	American Democracy	18
Hughes	Community Civics	10
	SUPPLEMENTARY TEXTS	
Magruder	American Government	16
Forman	American Democracy	13
Bryce	American Commonwealth	9
Ashley	New Civics	7
Hughes	Community Civics	7
	Total number of schools offering Civics	418

In connection with Civics current events are taught in about 85 percent of the schools. The course is considered unsuccessful in twelve schools and reported as not being liked by the students in seven schools. This means either that it is generally liked in at least five schools reporting it as unsuccessful, or that the course was not properly estimated. Two schools plan its discontinuance. Notebook work is done in 58 percent of the schools. The socialized recitation is found nearly as frequently. Supervised study is found in about one-third of the schools, and the stereopticon is used for teaching purposes in only 14 percent, this low percent being due to the nature of the subject matter.

Vocational Civics. This subject is found most frequently in the first year of the high school course, but it is offered occasionally in every other year. Usually it is restricted to pupils who are in the year for which it is scheduled. As with most of the other social subjects the textbook furnishes the subject matter for the course. The texts most frequently used are as follows:

PRINCIPAL TEXTS

Author	Title	Frequency
Gowin and Wheatly	Occupations	13
Giles	Vocational Civics	11
Hughes	Economic Civics	2

(No supplementary texts were reported)

Total number of schools offering Vocational Civics 38

The persons replying to the questionnaire feel that Vocational Civics is given more prominence than it should have. This may be inferred from the fact that it is required in 42 percent of the schools but that only 32 percent of the replies expressed the opinion that it should be required. Vocational Civics has been taught in sixteen schools for three years or longer, and in seventeen schools for a shorter period of time. It is reported as being unsuccessful in six schools, and whether or not this is due to the nature of the course, it is interesting to note that no school intends to discontinue this subject. Notebook work is required and the socialized recitation is found in approximately 50 percent of the schools that offer this course. The stereopticon is used in one-half that number and supervised study in about two-thirds of that number.

Economics. Economics is offered most frequently in the fourth or the third year of the high school, yet six schools offer it in the first year. The data indicate that 82 percent of these courses are based predominantly upon some textbook. The texts used most frequently for this purpose are as follows:

PRINCIPAL TEXTS

Author	Title	Frequency
Thompson	Elementary Economics	104
Ely and Wicker	Elementary Principles of Economics	37
Carver	Elementary Economics	32
Burch and Nearing	American Economic Life	27
Bullock	Elements of Economics	19
Marshall and Lyon	Our Economic Organization	17
	SUPPLEMENTARY TEXTS	
Bullock	Elements of Economics	8
Carver	Elementary Economics	7
Ely	Outlines of Economics	6
Marshall and Lyon	Our Economics Organization	6
Seager	Economics	6

Total number of schools teaching Economics 316

This subject has been offered in 208 schools for three or more years and in 69 schools for a shorter period of time. Reports indicate that in about 7 percent of the schools the course is disliked by the pupils. Thirteen schools report it as unsuccessful and four of these schools will not offer it again. Notebook work is required in 58 percent of the schools, the stereopticon used only in about one-fourth that number, and supervised study and the socialized recitation found in 30 percent and 47 percent respectively.

Sociology. Sociology appears most frequently in the fourth year of the high school, yet a large number of schools offer it during the third year. Eight schools teach it regularly to freshmen. As a rule, however, only upper classmen are permitted to enter the course. It alternates in most cases with Civics or Economics. As a rule the textbook is the basis of the subject. The texts that are most frequently used are as follows:

Author	*Title*	*Frequency*
	PRINCIPAL TEXTS	
Towne	Social Problems	67
Ellwood	Sociology and Modern Social Problems	46
Burch and Patterson	American Social Life	22
Tufts	Real Business of Living	4
	SUPPLEMENTARY TEXTS	
Ellwood	Sociology and Modern Social Problems	9
Towne	Social Problems	4
Burch and Patterson	American Social Life	4
Tufts	Real Business of Living	2
	Total number of schools teaching Sociology	156

Practically one-half of these schools teach current events in connection with this course. About 25 percent of the schools offering Sociology require it for graduation. The persons answering the questionnaire in about 50 percent of the schools believe the subject should be required. Hence, again theory and practice differ. Indications are clear that Sociology is fast coming into the high school curriculum. Its presence is reported by fifty-nine schools for three years or longer, and by seventy-five schools as having been introduced within the past three years. It is reported as unsuccessful in seven schools and as not being generally liked in three schools. It will not be offered next year by three schools. Notebook work is required in 44 percent of the

schools, the stereopticon is found in 16 percent, and supervised study and the socialized recitation are found in 35 percent and 57 percent, respectively.

Social Science. This course is offered most frequently to seniors or to sophomores. In nearly every instance it is limited to the students of the class for which it is offered. About three-fourths of the schools giving the subject adhere strictly to the textbook. The textbooks that are most frequently used are as follows:

PRINCIPAL TEXTS

Author	*Title*	*Frequency*
Hill	Community Life and Civic Problems	4
Burch and Patterson	American Social Life	4
Tufts	Real Business of Living	3
Towne	Social Problems	2
Total number of schools offering Social Science		31

Current events are taught in this course in about three-fourths of the schools. About 16 percent require the subject for graduation. Indications from the opinions of those persons answering the questionnaire are that the subject should be required much more frequently that it is at present. Social Science has been offered for three years or longer by seven schools, and for less than three years by nineteen schools. It is reported as not being generally liked in two schools, and as unsuccessful in one school which plans to discontinue it. Notebook work is required in 55 percent of the schools offering the subject, the stereopticon is used in 26 percent, supervised study in 32 percent, and the socialized recitation in 45 percent of these schools.

Educational Guidance. This subject is offered in a very few schools, and appears either in the first semester or in the last three semesters of the high school course. When offered in the first year it is likely that one phase of its work is related directly to the students' future work in the high school. This subject does not appear to have a very definite content, and probably the subject matter of the course usually depends upon the interests of the instructor. In several cases "How to Study" by Sandwick is mentioned as a basic text. Current events are taught in almost one-half of these courses.

Of the schools teaching Educational Guidance 20 percent require it for graduation. However, 56 percent of the persons who answered

the questionnaire think that it ought to be required. It is not generally liked in one of the five schools reporting its presence. Notebook work and the socialized recitation are used in 20 percent of these schools.

Vocational Guidance. This subject is offered in either the second or the first year of the high school course. In nearly every case the classes are restricted to students with standing equivalent to the year in which the course is scheduled. "Occupations" by Gowin and Wheatley is used almost universally as a text. No other text was mentioned more than once. One-third of the schools require this course for graduation but the conclusion based on the answers to the questionnaire is that it should be required to a much larger extent. This subject also has made its appearance within the last few years, although only nine schools report that it has been taught for less than three years. It is generally liked in all the schools represented, is reported successful, and will be offered again in every case. Notebook work is required in 50 percent of the schools. The stereopticon is used in 15 percent, and supervised study and the socialized recitation reported in 22 percent and 37 percent, respectively.

BULLETIN NO. 14

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE USE OF DIFFERENT TYPES OF THOUGHT QUESTIONS IN SECONDARY SCHOOLS AND THEIR RELATIVE DIFFICULTY FOR STUDENTS

by

WALTER S. MONROE,
Director, Bureau of Educational Research

and

RALPH E. CARTER
Associate Professor of Educational Psychology
Indiana University, Extension Division

PRICE 30 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

PREFACE

This bulletin is a report of one phase of a larger investigation relating to the study habits of school children. The types of questions asked by teachers of their students are important both because of the mental processes which occur in answering them and because an intimate relation exists between the questions asked and the detailed objectives toward which the students work.

Mr. Carter was consulted in the preparation of the questionnaire. He rendered valuable assistance in deciding upon the tentative list of types of thought questions. Later, advantage was taken of his presence at the University of Illinois during the Summer Session of 1922 to secure assistance in the preparation of this report. He is largely responsible for its general organization.

WALTER S. MONROE, *Director.*

February 1, 1923

The Use of Different Types of Thought Questions in Secondary Schools and Their Relative Difficulty for Students

Recent emphasis upon questioning in instruction. The first quantitative study of the questions which teachers use in the classroom was made by Miss Romiett Stevens.[1] In this analysis of questions a distinction was made between those that required memory only and those that required thought. Miss Stevens' discussion has been widely quoted and has been very influential in causing teachers to give more attention to the types of questions asked of their students. It is altogether likely that this investigation is primarily responsible for a material increase in the percent of thought questions which are asked in our schools today. Many of the most commonly used texts on methods of teaching devote separate space to the topic of questioning. Its importance has been especially emphasized by Charters,[2] Strayer,[3] Parker,[4] and Colvin.[5]

Analysis of thought questions. An examination of typical thought questions from the standpoint of the mental processes which they require in answering reveals certain significant differences. Some questions require the student to "compare" two or more ideas; others ask that he "summarize;" still others demand that he "give reasons why." The probable mental processes occurring in the reflective thinking called for rather than the form of the question or the wording of the answer have been made the basis of these types of thought

---

[1]Stevens, Romiett. "The question as a measure of efficiency in instruction," Teachers College Contributions to Education No. 48. New York: Teachers College, Columbia University, 1912.

[2]Charters, W W. Methods of Teaching. Chicago: Rowe Peterson and Company, 1912, p. 444. (Chapter XVII).

[3]Strayer, G. D. A Brief Course in the Teaching Process. New York: Macmillan Company, 1912, p. 315. (Chapter XI)

[4]Parker, S. C. Methods of Teaching in High Schools. Boston: Ginn and Company, 1915, p. XXV, 529. (Chapter XX)

[5]Colvin, S. S. Introduction to High School Teaching. New York: Macmillan Company, 1917, p. XXI, 451. (Chapter XV)

questions. Altho there is doubtless considerable overlapping in the types of mental processes which usually occur in answering such questions, there appears to be sufficient differences to justify the recognition of a number of types of questions. In this investigation the following tentative list of types of thought questions was used. A more refined analysis would yield probably some additional types, but, on the other hand, for practical purposes it is possible that some combinations of types might profitably be made. It is recognized that the mental processes of different students doubtless vary in answering the same question. In fact a thought question for one student may be a memory question for another student, or even for the same student on the following day.

1. **Selective recall**—basis given.

Name the presidents of the United States who had been in military life before their election.

What do New Zealand and Australia sell in Europe that may interfere with our market?

2. **Evaluating recall**—basis given.

Which do you consider the three most important American inventions in the nineteenth century from the standpoint of expansion and growth of transportation?

Name the three statesmen who have had the greatest influence on economic legislation in the United States.

3. **Comparison of two things**—on a single designated basis.

Compare Eliot and Thackeray in ability in character delineation.

Compare the armies of the North and South in the Civil War as to leadership.

4. **Comparison of two things**—in general.

Compare the early settlers of the Massachusetts Colony with those of the Virginia Colony.

Contrast the life of Silas Marner in Raveloe with his life in Lantern Yard.

5. **Decision**—for or against.

Whom do you admire more, Washington or Lincoln?

In which in your opinion can you do better, oral or written examinations?

6. **Causes or effects.**

Why has the Senate become a much more powerful body than the House of Representatives?

What caused Silas Marner to change from what he was in Lantern Yard to what he was in Raveloe?

7. **Explanation** of the use or exact meaning of some phrase or statement in a passage.

8. **Summary** of some unit of the text or of some article read.

9. **Analysis.** (The word itself is seldom involved in the question.)

What characteristics of Silas Marner make you understand why Raveloe people were suspicious of him?

Mention several qualities of leadership.

10. **Statement of relationships.**

Why is a knowledge of Botany helpful in studying agriculture?

Tell the relation of exercise to good health.

11. **Illustrations or examples** (your own) of principles in science, construction in language, etc.

12. **Classification.** (Usually the converse of No. 11).

What is the principle involved here? What is the construction? To what class or genus does this individual belong?

13. **Application** of rules or principles in new situations.

14. **Discussion.**

Discuss the Monroe Doctrine.

Discuss early American Literature.

15. **Statement of aim**—author's purpose in his selection or organization of material.

What was the purpose of introducing this incident?

Why did he discuss this before that?

16. **Criticism**—as to the adequacy, correctness, or relevancy of a printed statement, or a classmate's answer to a question on the lesson.

17. **Outline.**

18. **Reorganization of facts.** (A good type of review question to give training in organization.)

The student is asked for reports where facts from different organizations are arranged on an entirely new basis.

19. **Formulation of new questions**—Problems and questions raised.

What question came to your mind?

What else must be known in order to understand the matter under consideration?

20. **New methods of procedure.**

Suggest a plan for proving the truth or falsity of some hypothesis.

How would you change the plot in order to produce a certain different effect?

Relation to educational objectives of types of questions asked by teachers. Incidentally it may be noted that the types of questions used both for stimulating and directing the mental activity of the learner and for measuring the results of teaching reflect in a subtle way the educational objectives of a teacher. The objectives thus indicated may not agree with those stated by the teacher but they nevertheless are an index of the objectives toward which the students

direct their efforts. In studying, students tend to prepare to answer the kind of questions which they think the teacher will ask. It is altogether likely that the study objectives of students are influenced more by the kind of questions asked than by direct statements of aims in the course.

Purpose and method of this investigation. The purpose of this study, which is Sub-project II of our investigation of the study habits of high school pupils, was announced as follows: "To determine the extent of the use of different types of thought questions in actual school practise and the relative difficulty of these types for students." The method employed was to submit a questionnaire to a large number of teachers in the seventh and eighth grades and in the high school. In this questionnaire the list of the twenty types of thought questions given on pages 6-7 was reproduced. The teachers were urged to study these types carefully before answering any of the questions. It was pointed out that an attempt to classify a question according to this plan might lead to the conclusion that it is a combination of two or three types. However, in such a case, it is probably true that part of the work of answering has been done by the author in the textbook, leaving to the student only that phase of the question which would definitely classify it as belonging to one of the above types.

Plan of summarizing data. Usable replies were received from 199 teachers representing almost every school subject and all parts of the state. A few of these replies did not give answers to one or more of the questions but they were included in the tabulation for the other questions. By subject the number of replies were as follows: English, 41; History, 48; Science, 41; Mathematics, 31; Foreign Language, 12; Geography, 11; Commercial subjects, 6; Agriculture, 4. Taking only the replies from teachers of English, History, and Science there were 26 for grades VII and VIII combined; 34 for grades IX and X; and 45 for grades XI and XII. It was thought advisable to summarize separately the replies for English, History and Science. Those received from teachers of all other subjects were grouped together. The replies to Questions I and IV obviously do not lend themselves to statistical treatment. The data yielded by the other three questions are summarized in the following tables.

Limitations of replies to the questionnaire. Several teachers took occasion to mention difficulties which they experienced in filling out the questionnaire blank. That some difficulties were encountered

is not surprising because the concept of different types of thought questions was undoubtedly new to most, if not all, of the teachers. The limitations of space prevented an extensive description of each type. Undoubtedly teachers differed in their interpretations of the various types. Furthermore, in answering Question II they were asked not only to differentiate between thought questions and memory questions but also to classify the thought questions under the various types. In doing this, they had to rely upon their memory of the questions they had asked. For these reasons the replies to Question II must be thought of as giving only a very rough indication of the practise of teachers with reference to the types of questions which they asked. A few teachers reported that their replies were based on a careful analysis of lists of questions which have been used and preserved. If these lists were reasonably complete their replies should be more accurate than those received from teachers who relied upon their memory of the questions they had asked.

It may be pointed out that other methods of ascertaining the relative frequency of use of different types of thought questions would not eliminate all difficulties. For example, if a trained investigator should visit the classes of 199 teachers in various parts of the state a sufficient number of times to obtain a fair sample of the habitual practise and should try to make a record of all types of questions asked he would encounter difficulties in definitely classifying them. It is not possible to judge accurately of the mental processes stimulated by a question unless one is acquainted with the previous experiences of the pupils in the field of this question. This has been referred to already in pointing out that what constitutes a thought question for one pupil may be merely a memory question for another. The analysis of stenographic reports of lessons would present certain difficulties for the same reason. The teacher knows better than any one else what acquaintance a student has with a topic and what the textbook states about it. For this reason he has the advantage of a casual observer.

Additional types of questions mentioned. Question I was asked in order to obtain suggestions for supplementing the list of type questions for further study in this field and also to give each teacher an opportunity to report fully his own practise in the use of different types of thought questions in case he did not consider the list adequate. The replies to the questionnaire show that the number of teachers who did not report the use of all of the types given in the list

UNIVERSITY OF ILLINOIS

BUREAU OF EDUCATIONAL RESEARCH

COLLEGE OF EDUCATION

Types of Thought Questions

Name..............................Address..................................

School Subject...........................School grade......................

I. Write in on the blanks at the bottom of the list below any additional types of thought questions you use to any great extent in the grade and subject you have chosen. (Include these additional types in answering all other questions.)

II. What percent of *all* of your questions to the class are of each type listed? (See directions, p. 11). Put the percent of each type in the column on the right.

Types of Thought Questions	Percent of all questions
1. Selective recall—basis given............................	....
2. Evaluating recall—basis given............................	....
3. Comparison of two things—on a single designated basis. ...	....
4. Comparison of two things—in general....................	....
5. Decision—for or against..............................	....
6. Causes or effects....................................	....
7. Explanation..	....
8. Summary..	....
9. Analysis..	....
10. Statement of relationships............................	....
11. Illustration or examples..............................	....
12. Classification......................................	....
13. Application.......................................	....
14. Discussion..	....
15. Statement of aim....................................	....
16. Criticism...	....
17. Outline..	....
18. Reorganization of facts..............................	....
19. Formulation of new questions..........................	....
20. New methods of procedure.............................	....
21.	....
22.	....
23.	....
24.	....

III. In the left hand margin make an "X" opposite each of the five types for which students answers are least likely to be satisfactory. (See directions, p. 11)

IV. Choose three of the types you marked with an "X." State for each some of the students' most common faults in procedure in answering it. Mention faults of

omisson as well as faults of commission. Mention only faults peculiar to this type of question. Use other side of this sheet if necessary. (See directions below.)

V. If we consider all unsatisfactory answers made by students in school work, what percent of them, in your opinion, are due primarily to bad habits of procedure in answering questions? Give separate replies for memory questions ....% and thought questions............%

The following detailed directions for answering the questionnaire were given:

Before answering any of the questions, decide what subject and what grade (any from VII to XII) you will have in mind in all of your replies. Answer the questions with reference to only *one* school subject. Fill in the blanks at the top of the first page of the questionnaire.

Question I. Examine the list of types given above to see if there are not some other types of thought questions that you use to a considerable extent. Altho the given list may seem long and inclusive at first, it is not complete.

Question II. Before you put down your estimates for Question II, you should decide what is your proportion of thought questions and of pure memory questions. This question calls for a further analysis of your thought questions. It may be advisable to make rough estimates for all before you fill in the data on the questionnaire. After you have the differences between the types in mind, make the best estimate you can, even tho you may not be very sure in many cases. You are merely declaring your best judgment, not guaranteeing that it is infallible. *We want your opinion of what your practise is, not of what you think it ought to be.*

Question III. In Question III, we are concerned with the *process* of answering so we must assume that the student has the information necessary for a satisfactory answer if he will only use it as the type of the question requires. The common starting point for an informal diagnosis of a student's study faults is in his unsatisfactory answers. An unsatisfactory answer may indicate lack of information or certain faulty habits in answering questions. There are many cases in which teachers find out by other means that the student has enough information to give a satisfactory answer but has failed because he did not take the necessary steps and precautions in replying to the question. When some students correct examination papers returned to them, they are heard to say, "I knew that, but I didn't think of it," or "I knew that but I didn't know that was what the question meant." Some of the faults in procedure are common to all kinds of questions; others are peculiar to particular types. In Question III, you are asked to decide on the difficulty of the types for students. In other words, in which type do they have the worst habits of procedure?

Question IV. It will be helpful in answering Question IV to think of some particular questions of the type under consideration and then express the students' faults in fairly general terms.

Question V. As long as the unsatisfactory answer is a resultant of poor methods of preparation and poor habits of answering questions, we must correct the latter or take it into consideration in inferring what methods of study are needed. It is important to get an estimate from a large number of teachers of the percent of failures (in answers) that are due primarily to bad habits of procedure in answering questions.

is very much greater than the number mentioning additional types. Only thirty-nine teachers (19.6 percent) indicated that they used all twenty types of questions. Thirteen teachers supplemented the list, each writing in from one to three additional types. These teachers were distributed among the different subjects as follows: seven in History, three in English, two in Science, and one in Mathematics. Only four of those suggesting additional types of questions indicated that they used all twenty of the types given in the questionnaire list.

Some of the additional types of questions suggested are clearly included in the list given in the questionnaire. The following are typical: (A) "Connections between historical events," Type 4 or 6; (B) "Cause and results," included in Type 6; (C) "Application to the present time," included in Type 13; (D) "Determining the significant word in a statement," included in Type 19. Other suggested types are not so clearly included but several of them under a fairly loose interpretation may be placed with the types listed. Illustrations of these are: (A) "Imagining the results if conditions were different," a special case of Type 6; (B) "Questions on appreciation such as, which do you like best?" a fairly common form of Type 2; (C) "Identify a known character in a scene in the story when his name is not mentioned by the author," a special case of Type 12; (D) "What should you judge from these facts?" This last is broad enough to cover many types. Frequently, it would fall under Type 6. There were a few suggestive questions which are not so easily classified under the 20 types given. Some of these are: (A) "Trace the development," mentioned by two teachers; (B) "Estimate the importance of"; and (C) "Why is this statement true?" a very common question in geometry. By a very liberal interpretation these might be put under Types 9, 2, and 6 respectively but it is likely that many teachers would consider them sufficiently different to justify naming them as additional types.

The answers to the first question indicate that the list of types is reasonably complete. For practical purposes it is probably better to give a rather loose interpretation in classifying special cases than to try to extend the list. Largely for this reason the authors decided to limit the summary of the replies to the other parts of the questionnaire to the original twenty types.

Frequency of use of different types of questions. Question II of the questionnaire was asked in order to secure answers for the following:

1. Which types of questions are most commonly used by teachers in classroom instruction?
2. Which types of questions are most characteristic of the instruction in different subjects?
3. Do teachers in the more advanced grades tend to ask different types of questions from those in the lower grades?

Table I gives a detailed summary of the replies by forty-eight teachers of History. This table shows that four of these teachers did not report any use of questions requiring selective recall (Type 1); one teacher reported that this type of question formed 1 percent of all his questions; five teachers indicated that this type made up 2 percent of all the questions which they asked; eleven named 5 percent as an index of its frequency; eight teachers considered that 15 percent of their questions were of this type; and another eight teachers indicated that more than 15 percent of their questions called for selective recall. The median practise is 7.5 percent.

The outstanding characteristic of the table is the variability of practise which it indicates. In the case of each type of question there were one or more teachers who gave no indication of its use. For several of the types the number of teachers not indicating any use is surprisingly large. Furthermore, there is no type of question which did not receive an indication of at least 5 percent by one teacher. Most of the types received an indication of 10 percent or more. Thus, if we may assume that the replies to this portion of the questionnaire are a reasonably true indication of practise, it is clear that teachers of History vary widely in the types of thought questions which they ask of their pupils and hence necessarily vary widely in the detailed objectives which their pupils strive to attain.

Similar tabulations were made for English, Science and other subjects combined. The median frequencies for the four groups of subjects are given in Table II. If we consider only English, History and Science, students are most commonly required to answer the following types of thought questions: cause and effect, Type 6; selective recall, Type 1; discussion, Type 14; and evaluating recall, Type 2. The types of questions which are asked least frequently are: formulation of new questions, Type 19; new methods of procedure, Type 20; and reorganization of facts, Type 18.

Different subjects require different mental processes. An examination of Table II reveals that the frequency of use of some of the

TABLE I. TYPES OF THOUGHT QUESTIONS ASKED BY 48 TEACHERS OF HISTORY

Types of Questions	Percent of all Thought Questions																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Over 15	Median
1. Selective recall	4	1	5	2		11		2		1	6					8	8	7.5
2. Evaluating recall	2	1	4	4	3	7	2	3	3	2	5		3			6	3	7.3
3. Comparison—single	6	5	4	1	4	16			2	2	5		1		1	1		5.3
4. Comparison—general	4	1	8	3	2	13	2		4	1	6					2		5.5
5. Decision	10	11	5	4	3	10		1	1		2					1		2.6
6. Causes or effects	1		1	1	3	6	2	2	3		12		1			4	10	10.4
7. Explanation	15	4	8	8	1	5	1				3					2		2.6
8. Summary	5	1	7	1	2	15			1	1	10					1	3	5.5
9. Analysis	7	7	8	4	5	11		1	1	1	2							3.5
10. Relationships	13	4	7	4	5	4	4	1	4		2							3.0
11. Illustration or examples	25	5	3	2	4	6					1		1			1		1.0
12. Classification	26	4	6	2	2	6	1		1								1	.9
13. Application	24	5	7	3	3	2		1			2		1					1.0
14. Discussion	4		1		1	6	2	3	3	1	12					6	9	10.3
15. Statement of aim	24	7	6	3	4	2										1		1.0
16. Criticism	12	5	12	4	4	6	1	1	1		2							2.6
17. Outline	6	3	5	4	4	11	1				9				2		1	5.2
18. Reorganization of facts	14	8	8	6		5			2		4					1		2.3
19. Formulation of new questions	31	3	3	7	3				1		1							.8
20. New methods	24	5	9	2	3	3			1		1							1.0

TABLE II. MEDIAN FREQUENCY OF USE OF DIFFERENT TYPES OF THOUGHT QUESTIONS IN ENGLISH, HISTORY, SCIENCE AND OTHER SUBJECTS

Types of Questions	Median percent of Use			
	English	History	Science	Others
1. Selective recall	5.4	7.5	5.3	5.1
2. Evaluating recall	5.9	7.3	2.8	.96
3. Comparison—single	5.4	5.3	3.9	1.6
4. Comparison—general	5.0	5.5	3.1	1.9
5. Decision—for or against	4.4	2.6	1.1	1.2
6. Causes or effects	5.8	10.4	10.2	4.8
7. Explanation	10.0	2.6	3.3	5.1
8. Summary	5.6	5.5	2.9	2.1
9. Analysis	5.7	3.5	1.8	3.1
10. Relationships	1.8	3.0	5.7	3.4
11. Illustration or examples	3.6	.96	8.6	5.1
12. Classification	1.6	.9	5.2	5.1
13. Application	2.9	1.0	5.1	5.5
14. Discussion	5.7	10.3	5.3	1.1
15. Statement of aim	4.4	1.0	.7	.8
16. Criticism	5.2	2.6	2.9	3.9
17. Outline	3.8	5.2	.98	.8
18. Reorganization of facts	2.1	2.3	.98	.9
19. Formulation of new questions	1.9	.8	1.7	1.8
20. New methods	2.1	1.0	1.5	3.6

types of thought questions varies in the different subjects. For example, Type 7 "Explanation of the use or exact meaning of some phrase or statement in a passage" is the most frequently used type of question in English but it is eleventh in frequency of use in History. Asking students to give an illustration or example, Type 11, is next to the most frequent type of question in Science but it is seldom used in History. Type 10, which asks the student to state a relationship, is frequently used in Science but is infrequently asked of students in English. Summaries and outlines are required more frequently in History and English than in the other subjects. Thus, if our assumption that different types of questions require different mental processes is valid, we have here evidence that different types of mental processes are required of students in the different subjects. Hence, we should expect to find that the problem of directing the learning of students is different in different subjects.

TABLE III. PERCENT OF TEACHERS REPORTING NO USE OF CERTAIN TYPES OF THOUGHT QUESTIONS

Types of Questions	Grades VII-VIII	Grades IX-X	Grades XI-XII
1. Selective recall	3	18	11
2. Evaluating recall	6	29	13
3. Comparison—single	18	9	18
4. Comparison—general	9	18	13
5. Decision—for or against	21	32	27
6. Causes or effects	9	6	4
7. Explanation	23	26	18
8. Summary	6	23	22
9. Analysis	21	23	29
10. Relationships	35	18	20
11. Illustration or examples	41	26	16
12. Classification	50	29	24
13. Application	46	35	22
14. Discussion	9	15	13
15. Statement of aim	63	46	42
16. Criticism	26	23	24
17. Outline	26	21	31
18. Reorganization of facts	35	41	42
19. Formulation of new questions	68	44	44
20. New methods	44	53	32

Variations in the use of type questions in different school grades. The replies of 113 teachers have been summarized in Table III on the basis of the school grades in which they were teaching. The subjects involved are English, History, Science and Geography. Instead of giving the median frequency of use, the percent of teachers indicating no use of the given type of question has been calculated. Thus, 3 percent of the teachers in the seventh and eighth grades reported no use of selective recall. In the ninth and tenth grades 18 percent of the teachers did not use this type and in the eleventh and twelfth grades, 11 percent. The general impression prevails that students in advanced classes are asked more difficult types of questions than students in the lower grades. A comparison of the data for the three grade groups shows that the differences are not very great. In general, it appears that the variations in the use of thought questions are greater for different subjects than for the different grades.

The relative difficulty of the different types of questions for students. In Question III teachers were asked to designate the five

TABLE IV. TYPES OF QUESTIONS FOR WHICH STUDENTS' ANSWERS ARE LEAST SATISFACTORY

Types of Questions	English	History	Science	Others	Total	Rank
1. Selective recall........	5	10	8	14	37	15
2. Evaluating recall......	12	12	11	10	45	8
3. Comparison—single....	6	7	9	5	27	19
4. Comparison—general..	12	12	9	14	47	6
5. Decision—for or against	1	9	2	6	18	20
6. Causes or effects.......	17	17	19	24	77	1
7. Explanation..........	15	12	15	25	67	3
8. Summary.............	10	14	9	14	47	6
9. Analysis...............	13	15	11	21	60	5
10. Relationships.........	5	17	10	20	42	10
11. Illustration or examples	5	4	10	14	33	17
12. Classification..........	5	4	9	16	34	16
13. Application...........	16	12	18	29	77	1
14. Discussion............	10	14	5	10	39	13
15. Statement of aim......	15	12	5	10	42	10
16. Criticism.............	8	14	7	10	39	13
17. Outline...............	6	12	6	9	33	17
18. Reorganization of facts.	12	19	13	19	63	4
19. Formulation of new questions...........	9	6	8	18	41	12
20. New methods.........	6	7	7	24	44	9

types of questions in which "students' answers were least likely to be satisfactory." A summary of their replies is given in Table IV. The first line of this table should read as follows: Questions calling for selective recall were named among the five most difficult types by five teachers of English, ten teachers of History, eight teachers of Science, and fourteen teachers of other subjects. This type of question ranks fifteenth in the frequency of mention among the five most difficult types. The two types most frequently mentioned were Type 6, causes and effects, and Type 13, application. Both of these were mentioned by seventy-seven teachers or 38.7 percent. Altho we may say that in the judgments of those answering the questionnaire these two types are the most difficult for students, it should be noted that over 60 percent of the teachers did not list them among the five most difficult types.

Here, as in other tables, the variation in the replies of teachers is very extreme. Even when we consider the teachers of a single subject we find marked variations in their judgments concerning

the difficulty of the different types of questions. No type of question is so easy that it is not included among the most difficult five by several teachers. This variation in judgment is probably due largely to the fact that teachers have not recognized the distinction in types of questions which are made here and, furthermore, they have not analyzed the responses of their students in order to ascertain the nature of the difficulties which the students encounter in answering the questions.

Common faults of procedure in answering different types of questions. In Question IV the teachers were asked to select three of the types which they considered most difficult and to state for each of these some of the students' most common faults in procedure in answering them. An examination of the replies to this question suggests that a considerable number of teachers are satisfied with telling students that the answer is wrong, or what the answer should have been, and that they fail to give much thought to the students' faults of procedure in answering the question. It may be pointed out that questions are only a means to an end and this end is the education of the student. The answer is in itself relatively unimportant. The question fulfils its function only when it stimulates educative processes in the mind of the student. If the mental processes which the question initiates are not educative the asking of the question has been largely futile. Hence, it becomes highly important for the teacher to give attention to the procedure which the student employs in answering questions in order that the faults of procedure may be corrected so that the student's mental processes will become most effective in his education.

Of the 199 teachers from whom replies to the questionnaire were received, twenty-three did not attempt to answer Question IV at all and twenty-four others discussed all three types of questions together or stopped after discussing only one or two. Furthermore, a number of other teachers gave answers which show that they failed to grasp the significance of this question. In substance they said that the difficulty in students' procedure in answering a question is inability to do what is called for. For example, one said that the fault in answering Type 2, evaluating recall, is that "pupils are not always able to evaluate." Another said that in questions on aim "they seem unable to see any motive behind the statement of the author." Still another said that the fault in Type 6, causes and effects, is that "they are unable to see causes and effects." Still others mentioned objective faults in the

answers rather than faults in procedure. For example, one teacher mentioned that "minor details are included while leading points are omitted." Another stated that "applications are not as good as could be expected." Still another teacher simply stated that "the answers are inaccurate." General faults, such as "lack of practise in answering questions," "carelessness," "lack of concentration," "pupils do not know how to study," and so on were mentioned by one or more teachers.

It may be pointed out that teachers who fail to become definitely conscious of the difficulties which their students encounter will be unable to give them much definite constructive assistance in this phase of their learning. Probably the most significant conclusion to be drawn from the answers to this part of the questionnaire is that many teachers are failing to give attention to the procedure which students use in answering questions. They appear to be concerned largely with the accuracy of the answer and when it is wrong they fail to seek the cause in the procedure which the student has used.

Several of the faults which appear to be suggestive are given below. This list is not a complete statement of the faults of students in answering questions but should prove helpful to teachers who are desirous of ascertaining the reasons why students fail to answer questions satisfactorily.

Type 2.

The pupil fails to see real basis on which evaluation should be made.

Type 3.

A student seldom stays "on a single designated basis."

Type 3 or 4.

The pupil begins to answer the question before he has thought out what points should be included in the comparison. He often includes minor comparisons and omits the important points.

Unless there is a single designated basis for comparison, pupils fail to see all of the possibilities of comparison. The slothful pupil is content with one or two obvious comparisons.

The pupil can state some information about each thing compared, but omits the comparative connection between the two things. The pupil gives comments on one thing and omits the discussion on the other thing, showing how the two are alike comparatively.

In making comparisons, students usually give characteristics of things to be compared, or describe them separately but fail to make a complete comparison.

Type 4.

In comparing two things in general the worst fault is in the students' forgetting some of the important points to be compared. Often enough suggestions must be given so that the question finally should be placed under Type 3.

Pupils frequently fail to distinguish the important from the trival. The pupil's prejudices—likes and dislikes—influence the comparison.

Type 6.

Students confuse cause and effect. Method requires real *thinking and reasoning*—the hardest thing any student can be asked to do.

Students often know a statement is true but on account of lack of thoroness can not give causes or reasons.

Type 6 or 7.

Pupils, when asked, "Why?" seem to forget all reasoning and usually state the first thing that comes to mind and fail also to go far enough in explanation.

Type 7.

Pupils have difficulty in finding words to express the thought without repeating word for word parts of the passage to be explained. The chief difficulty seems to be the limitations of their vocabulary.

The trouble here is that pupils seem to have failed to study the statement under question sufficiently and so have quite an indefinite notion of the words used. And too, even if they succeed in getting a fairly good understanding of the text read, they are too careless of the choice of their own words in reproducing the meaning exactly. They are too hasty in the study of the passage and have not yet been trained to go slowly and get facts one at a time.

Type 8.

A discussion, like a summary often lacks point. The student fails to organize material and grasp essentials.

Students fail to properly organize data mentally before expressing them.

Type 9.

Here pupils are often prone to feel satisfied if only a few of the factors or qualities involved in the analysis are learned and reproduced. Many times but one factor will be offered when to completely analyze the problem will require several. Being allowed to "get off" with a minimum of thinking is one of the chief causes of poor analysis.

Type 10.

Pupils often fail to get the facts correlated before they attempt to state the existing relationships.

Pupils do not stop to analyze in order to see the relationship but answer without giving much thought.

Pupils answer various separate facts instead of showing relationship. These facts are usually true and have a bearing upon the question but the actual relationship is usually missing in the answer. That is, the pupil does not seem to see how a change in one fact would influence some other related fact.

Type 11.

Illustrations or examples. This is not so difficult after the student has had some practise, but at first he always tries to find his illustrations in some field remote from his own. It takes him a while to realize that scores of illustrations come daily under his observation—*if he observes.*

Type 12.

In classifying, the pupils make their answers too readily without first trying one class then another or without thinking whether the one they choose actually does fit better and why.

The student's failure in Type 12 is due mainly to his failure in Type 7. He fails to decide correctly upon a construction because he has not thought out carefully the exact meaning of the sentence. This in turn is due to mental laziness or carelessness.

Type 14.

He discusses each phase at random and separately without relating it to others.

A discussion like a summary often lacks point. The student fails to organize material and grasp essentials.

Students fail to properly organize data mentally before expressing them.

Students do not group their facts together sufficiently to discuss them properly.

Type 17.

A pupil doesn't really outline into main facts with heads under them, but makes long statements one after another, or puts steps in before their time.

General.

Pupils will not read and get a thoro knowledge of the subject-matter first.

With Types 8, 14, and 17 pupils do not summarize or discuss or outline in a way to make clear to another the things under discussion. They are not logical in the arrangement of their ideas. They answer, at least 85 percent of them, only well enough so that one who knows already can tell that they know, but only about 15 percent can arrange and subordinate well enough to be clear to the unintiated.

A common fault is careless reading of the question (or hearing it as the case may be). This is especially true in Algebra. Another fault is "saying the first thing that comes into mind" instead of carefully considering its application to the problem under discussion. In Geometry, for example, the student does not rapidly recall all theorems pertaining to the subject in hand, eliminate those inapplicable, and so discover the correct one.

Percent of unsatisfactory answers due to faulty procedure in answering questions. In Question V the teachers were requested to indicate separately for memory questions and for thought questions the percent of unsatisfactory answers which, in their opinion, were due primarily to bad habits of procedure. A summary of their replies is given in Table V. As previously indicated it appears certain that a considerable number of teachers have given relatively little thought to the faults in the procedure used by students in answering questions. Consequently, the replies to this question must be considered as representing no more than rough estimates. In the case of thought questions the median percent of unsatisfactory answers due to faulty procedure is approximately 50 for all of the subjects. In the case of memory questions faulty procedure is considered to be a less potent cause. There are two factors that probably tend to

TABLE V. PERCENT OF SATISFACTORY ANSWERS DUE TO FAULTY PROCEDURE IN ANSWERING QUESTIONS

(As Contrasted with Lack of Information)

Percent	Thought					Memory				
	English	History	Science	Others	Total	English	History	Science	Others	Total
95	1	1			2	1	5			1
90	3	1	2		6	2		2		4
85			3	3	6	1			1	2
80	2	2		5	9	2				2
75	5	7	2	13	27	4	1	4	2	11
70			1	5	6			1	1	2
65	2		2		4					0
60	3	5	3	6	17	4	1		5	10
55					0					0
50	3	7	8	10	28	3	5	5	12	25
45	1				1					0
40		7	4	3	14	3	6	1	7	17
35	3	1	4		8		1	1	1	3
30	3	2	3	9	17				6	6
25	4	7		6	17	6	14	11	11	42
20		3	7	4	14	7	5	2	6	20
15					0	1	3	1	2	7
10	7				7	2	4	8	7	21
5	1	1		1	3	1	3	3	4	11
0				1	1					0
Median	48	48	48	51	50	40	25	25	25	26

make these estimates lower than they should be. Many teachers showed by their replies to Question IV that they had very little experience in analyzing answers for the purpose of determining the cause of faulty answers. Naturally they could not be expected to recognize the real extent of such faults of procedure on the part of the members of their classes. Other teachers use such a limited range in types of questions that they would not have a chance to observe as many faults in answer technique as teachers who use a greater variety of types of questions.

Relation between frequency of use and difficulty. There seems to be very little relation between the difficulty of a type and the frequency of its use. Type 6, causes and effects, is the most commonly used type and at the same time it, together with Type 13, application, is considered the most difficult for students. Only one of the five types reported as most difficult, Type 18, reorganization of facts, is among the five least frequently used; only one of the five types reported as least difficult, Type 13, comparison on a single basis, is among the five most frequently used.

Relation of frequency of use of types of questions to emphasis upon different types of learning. An earlier bulletin[1] of the Bureau of Educational Research reports the relative emphasis on different types of textbook study in the various subjects. We now have some data on the question, "How does the frequency of use of types of questions correspond with the emphasis on the different types of textbook study required of students?" Some of the types of study recognized in the report are too broad to be tested by any one type of thought question. There are, however, question types that correspond rather closely with certain of the study types.

To get a rough measure of the relation of the frequency of use of the different types of thought questions to the emphasis upon the types of learning, we may arrange the types in order of frequency of use and divide the list into quartiles (four equal parts). For instance, in English, Study Type II is in the first quartile (i. e. the three most frequently required study types); Study Type X is in the fourth

[1]Monroe, Walter S. "Types of learning required of pupils in the seventh and eighth grades and in the high school." University of Illinois Bulletin, Vol. 19, No. 15, Bureau of Educational Research Bulletin, No. 7, Urbana: University of Illinois, 1921, 16. p.

Types of Textbook Study	Types of Questions
II. Summary of central ideas in the lesson studied.	8. Summary.
III. Prepare an outline of principal points and supporting details in the lesson arranged to show order of relative importance and relationship.	17. Outline.
VI. Discovery of collateral or illustrative material for topics or problem under discussion.	11. Illustration or examples.
VIII. Appreciation of the significance of each word used in a concisely expressed statement or principle.	7. Explanation of the use or meaning of some phrase or statement in a passage.
IX. A clear comprehension of the essential conditions of a problem which is to be solved.	9. Analysis.
X. Discovery of new or supplementary problems related to the topic being studied.	19. Formulation of new questions.

quartile (i. e. one of the least frequently required study types).[1] Table VI shows the extent of agreement in rank in frequency of use of some corresponding types of study and questions. The first column gives by number the corresponding types as listed above. The first number given in the columns for the separate subjects designates the quartile rank of the study type on the same horizontal line; the second number designates the rank of the corresponding type of question. It will be noted that the rank agrees in seven out of eighteen possible cases. In English the disagreement is greatest in the case of questions calling for analysis (first rank) and study requirements of a clear comprehension of the conditions of a problem (fourth rank); in History, the discovery of collateral or illustrative material ranks second in study requirements but questions calling for illustrations or examples rank fourth in use; in Science, the same study requirement ranks fourth and the corresponding question type ranks first. Thirty-three percent of the Physics and Chemistry teachers in the first questionnaire mentioned Study Type X, discovery of new or supplementary problems, as a major type of study in these subjects, yet 44 percent of the teachers of these subjects by their admission do not ask any questions of Type 19, formulation of new questions, to their

[1]These facts are derived from Table II of the Bulletin referred to above.

TABLE VI. CORRESPONDENCE OF TYPES OF STUDY AND TYPES OF QUESTIONS

Corresponding Types		English		History		Science		Agree	Vary	Vary Widely
Study	Quest.	Study Rank	Quest. Rank	Study Rank	Quest. Rank	Study Rank	Quest. Rank			
II	13	1	2	2	2	3	3	Hist.+Sci.	English	
III	17	2	3	1	2	4	4	Sci.	Eng.+Hist.	
VI	11	3	3	2	4	4	1	Eng.		Hist.+Sci.
VIII	7	2	1	3	3	3	2	Hist.	Eng.+Sci.	
IX	9	4	1	4	2	3	3	Sci.		Eng.+Hist.
X	19	4	4	3	4	2	3	Eng.	Hist.+Sci.	

high school juniors and seniors. It would seem from the data that some teachers may not be using the types of questions that would stimulate and test students' achievements in some of the types of study they are expected to use.

Conclusions. Probably the most significant conclusions to be drawn from this investigation are: (1) teachers are not sufficiently conscious of the types of questions which they are accustomed to ask and of the significance of these types, and (2) in general teachers do not analyze unsatisfactory answers to questions in order to ascertain whether such answers are due to a faulty technique on the part of the student.

A number of other conclusions are worthy of mention. (1) Teachers of the same subject vary widely in the extent of their use of different types of thought questions. (2) The frequency of the use of a given type does not seem to depend very much on (a) the school grade, (b) the subject, or (c) the supposed difficulty of the type. It is very likely that some teachers, who expect their students to use certain types of study, do not use the types of questions that are best suited to test their students' achievements due to those particular types of study. (3) Teachers individually think that certain types of questions are more difficult for students than others. However, there is no very great agreement among teachers as to the relative difficulty of the various types.

BULLETIN NO. 15

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE CONSTANT AND VARIABLE ERRORS OF EDUCATIONAL MEASUREMENTS

by

WALTER S. MONROE
Director, Bureau of Educational Research

PRICE 25 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

TABLE OF CONTENTS

PREFACE

This bulletin, based upon a number of investigations conducted by the Bureau of Educational Research, brings together data relating to the errors encountered in educational measurements. Its purpose is to call the attention of users of educational tests to the nature and magnitude of the errors which they will encounter. The bulletin is not intended as a criticism of educational tests, but rather as an aid to a more intelligent use of them.

WALTER S. MONROE, *Director.*

October 4, 1923

THE CONSTANT AND VARIABLE ERRORS OF EDUCATIONAL MEASUREMENTS

CHAPTER I

INTRODUCTION

Educational measurements for many generations have been made by means of written examinations and teachers' estimates. These, we have been told, are subject to very large errors. Standardized educational tests have been proposed as instruments for obtaining more accurate measures. These instruments, however, do not yield measures involving negligible errors. In our measurements of ability in reading, spelling, arithmetic and other school subjects, we have not and are not likely to approximate the accuracy and precision with which the scientist is able to measure height, volume, temperature, and mass. We obtain, in fact, errors very much greater than those with which we deal in ordinary physical measurements.

In order to avoid misleading interpretations of educational measurements, it is necesary for us to be familiar with the nature and significance of the errors which we encounter. We need also to have some concept of the absolute magnitude of these errors. In this bulletin we attempt to answer the following four questions with reference to the errors encountered in the measurement of achievement and general intelligence by means of standardized educational tests:

1. What are the causes that tend to produce the errors encountered in educational measurements?
2. What is the nature of these errors?
3. What is the magnitude of the errors to be expected?
4. What is the effect of these errors upon the average, standard deviation and coefficient of correlation?

Variations in testing conditions tend to produce errors in educational measurements. A pupil's performance and hence his score on an educational test depend upon a number of factors. For example, it has been found that recent instruction may operate to increase the scores of pupils. Impending school events or other distractions may tend to lower their scores. Among the factors which

may be easily specified as influencing a pupil's score are the following: his emotional status, his physical condition, the effort which he makes, the set of his mind, the recency of instruction in the field of the test, the acquaintance which he has with the type of exercise he is asked to do, the manner in which the test is presented, the particular directions which are given him with reference to methods of work, the distractions, and the time allowed. There are other factors, such as the attitude of the teacher toward the test, which are more subtle in character but which doubtless in many cases operate to increase or decrease the scores of many or all of the pupils tested.

A test is standardized with reference to certain specific testing conditions. The use of a standardized educational test implies that these same testing conditions are to prevail when it is given to a group of pupils. This means that standard testing conditions must be secured for each pupil as well as for the group as a whole. If this is not done the norms do not constitute a valid basis for interpreting the scores. Any variations from the standard testing conditions tend to produce variations in the performances of some or all of the pupils. These variations constitute errors of measurement.

Constant and variable errors of measurement. Errors of measurement are of two types (1) constant errors and (2) variable errors.

A constant error is one which has the same magnitude for all of the scores of a given group. In other words the presence of a constant error results in all the scores of this group being either too high or too low. In the field of physical measurement we have an illustration of a constant error when a merchant gives short weights, such as a grocer who uses a peck or bushel measure which has a false bottom. The group of scores to which a given constant error applies may be those of a class, a school system or a group of school systems. It is possible that a given constant error would affect the scores made by boys and not those made by girls even when both sexes are tested together. Furthermore, it should be noted that a constant error may be either positive or negative.

A variable error of measurement is one which varies or differs in magnitude for the several scores of a given group. We may secure an illustration of variable errors in the field of physical

measurement by having a group of persons guess the length of a given object, for example, a table or even a pencil. If these guesses are made independently they are found to extend over a considerable range. In order to determine the magnitude of the variable errors involved in any guess it is necessary to determine the true length. In our illustration this might be done by having the length carefully measured by means of a yardstick or a tapeline. However, if we have secured a reasonably large number of guesses we may obtain an approximately true measure of the length by taking the average of the guesses. The difference between the true measure and any guess constitutes a variable error.[1] Some of these differences are positive and some negative. A few approximate zero.

In the field of mental measurements it is generally not possible to obtain true scores. Hence, we cannot calculate the magnitude of the variable error in a given score, but the concept of the true score aids us in understanding the nature of the variable errors of measurement. Approximately half of the variable errors for a given group of scores are positive and approximately half negative. If they were assembled for a frequency distribution the shape would approximate the normal probability curve with the average at zero. For a few measures the variable error would be relatively large, either positive or negative, but most of them would be near zero.

Constant and variable errors of measurement occur simultaneously. The situation may be represented by the following equation:

Obtained score = true score + constant error + variable error.

In this equation both errors may be either positive or negative, or one positive and the other negative. However, a constant error will have the same sign for all members of a group, that is, if it is positive for one pupil it will be positive for all of the pupils. Variable errors change signs within the group. Altho the two errors occur simultaneously, it is helpful to consider them separately and to treat each independently of the other.

---

[1]This statement is not strictly accurate. As we shall point out in a later paragraph, constant errors and variable errors occur at the same time. Thus, the difference may be the algebraic sum of the constant error and the variable error. However, in our illustration from the field of physical measurements, it is unlikely that there will be a large constant error. In the field of mental measurements there may be a relatively large constant error.

CHAPTER II

CAUSES, NATURE, AND MAGNITUDE OF CONSTANT ERRORS

Evidence of the presence of constant errors in educational measurements. 1. Constant errors due to acquaintance with the test. It is obvious that if a test is new to a given group of pupils, one significant change in the testing conditions attends its second administration. The test is no longer new to the pupils even if a duplicate form is used. When it is given a third or fourth time there is an added acquaintance with the type of exercise and the general form of the test. The taking of a test in itself thus introduces a change in the testing conditions which can not be eliminated. In order to secure evidence of the constant error due to the effect of acquaintance with a test it is necessary only to give the test twice to the same group of pupils under testing conditions which otherwise are as nearly the same as possible and to compare the averages of the two sets of scores. The difference between the average of the first trial scores and the average of the second trial scores is an index of the magnitude of the constant error resulting from the change in the testing conditions due to the pupils' acquaintance with the test. This difference, however, should not be interpreted as being the true magnitude of the constant error of the second trial scores. It is possible that the first trial scores also involved a constant error due to failure to secure standard testing conditions. However, when the averages of the two sets of scores are not equal we have evidence of the presence of a constant error and an indication of its magnitude.[1]

The Illinois General Intelligence Scale[2] was given twice to several hundred pupils in Grades III to VIII inclusive. After making due allowance for the inequality of the two forms of this scale[3] the

[1]In case different forms of a test are used in the two applications, it will be necessary to inquire concerning their equivalence and to make an appropriate allowance for any lack of equivalence in comparing the two averages.

[2]Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 69.

[3]See page 10 of the bulletin just referred to.

difference between the averages of the two sets of scores was approximately five points, or six months of mental age. In the eighth grade in which unusual testing conditions appear to have prevailed the difference was considerably greater. For Monroe's General Survey Scale in Arithmetic the difference between the average of the first trial scores and that of the second trial scores was approximately 3.2 points in Grades III to V, and 4.5 points in Grades VI to VIII. The writer recently had two forms of the Thorndike-McCall Reading Scale given to several groups of pupils. The average of the first trial scores (Form 2) was 47.78, the average of the second trial scores (Form 3), 51.69. Investigation has shown that these two forms are approximately equivalent. Hence, the difference between these two average scores indicates the magnitude of the constant error introduced by acquaintance with the form of the test.

One investigator[4] has reported data which is evidence of the presence of a constant error in the second trial scores yielded by the Burgess Picture Supplement Scale for Measuring Silent Reading Ability. Form 2 was given on the day following that on which Form 1 was used and care was exercised to secure as nearly the same testing conditions as possible. After discarding the records of all pupils who did not take both forms of the test the median scores for the two forms were 43 and 64. A part of the difference between these two median scores is undoubtedly due to the inequality of the two forms of the test used. This is shown by the fact that when the order of giving the two forms was reversed with another group of pupils, the median score dropped from 55 to 49. In a third group where Form 1 was given a few minutes after Form 2, the median score dropped from 58 to 49.

2. **Evidence of constant errors introduced by lack of equivalence of duplicate forms of a test.**[5] Altho the duplicate forms of a test are generally constructed so that they are expected to yield equivalent scores and to be used interchangeably, experience has shown that these forms are not always equivalent. Evidence of a

[4]Daley, H. C. "Equivalence of Forms 1 and 2 of the Burgess Picture Supplement Scale for Measuring Silent Reading Ability," Journal of Educational Research, 4:71, June, 1921.

[5]The lack of equivalence of duplicate forms of a test results in a constant error only when it is neglected as is the case when the same norms are used for interpreting the scores yielded by both forms or when comparisons are made between the scores yielded by the different forms without making due allowance.

TABLE I. DATA SHOWING THE EQUIVALENCE OF DUPLICATE FORMS OF TWO SILENT READING TESTS

Burgess Picture Supplement Scale			Thorndike-McCall Reading Scale		
Score	*Form 2*	*Form 3*	*Score*	*Form 2*	*Form 3*
20	1		66		6
19	1	1	63	3	9
18	6	1	60	5	9
17	7		57	17	19
16	7	4	54	14	37
15	4	5	51	46	18
14	9	4	48	36	65
13	12	9	45	99	72
12	18	18	42	81	61
11	20	15	39	43	60
10	27	21	36	44	49
9	30	23	33	27	20
8	26	33	30	6	4
7	17	31	27	6	2
6	19	25	24	1	
5	17	21	21	1	1
4	13	13			
3	9	19			
2	7	5			
1	7	6			
0	5	9			
Total.........	262	263		429	432
Median.......	9.37	8.08		45.18	45.78
Average.......	8.77	7.61		44.01	44.84

constant error due to such lack of equivalence is furnished by the illustration given in the preceding paragraph. More exact evidence may be secured by arranging the duplicate forms in alternate order, and distributing them in this manner to pupils as they happen to be seated in the classroom. Thus, if Form 1 and Form 2 are being compared the first, third and fifth pupils will have a copy of Form 1, the second, fourth, sixth, etc., of Form 2. If it is decided to secure information for three forms of one test at a time, a similar arrangement will result in every third pupil having a copy of the same form. Form 2 and Form 3 of the Thorndike-McCall Reading Scale were arranged in this way and given to several hundred children. The same procedure was followed with reference to the Burgess Picture Supplement Scale for Measuring Silent Reading Ability. The distribution of scores from the different forms of the two tests is given in Table I. Both the median and the average scores for the Thorndike-McCall Reading Scale show that Form 2 and Form 3 are approximately equivalent. In the case of the Burgess Picture

Supplement Scale, the differences between the medians and the averages indicate a lack of equivalence which can not be neglected safely when precise comparisons are being made between scores yielded by the two forms.

By a similar method the equivalence of the duplicate forms of the Illinois General Intelligence Scale, Monroe's General Survey Scale in Arithmetic and Monroe's Standardized Silent Reading Tests, Revised, was studied.[6] The evidence collected for these three measuring instruments indicates that the different forms are slightly lacking in equivalence. This is especially true of the measures of rate yielded by the silent reading test. Thus, it has been considered necessary to give correction numbers whereby the scores yielded by one form of the test may be reduced to a basis comparable with those yielded by the other forms.

A similar study of the three forms of Monroe's Standardized Silent Reading Tests indicated a marked lack of equivalence.[7] In order to eliminate the constant error due to this cause corrections have been calculated which may be used to reduce the scores yielded by the different forms to a comparable basis. Separate sets of norms have been stated for each form.

3. **Evidence of constant errors due to instruction functioning as coaching.** When any considerable period of time elapses between two trials on a given test, the instruction which pupils receive during this interim may materially influence their second trial scores. In a recent investigation[8] by the Bureau of Educational Research it was found that for a group of 134 children the increase of the second trial scores on the Illinois General Intelligence Scale over the first trial scores was equivalent to slightly more than four years in mental age. The two trials were six months apart and hence the normal increase to be expected would be six months. If we assume that the first trial scores were accurate, it follows that the constant error introduced in the second trial scores was in the neighborhood of three

[6]Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin, No. 6. Urbana: University of Illinois, 1921, 70 p.

[7]Monroe, Walter S. "Report of Division of Educational Tests for '19-20." University of Illinois Bulletin, Vol. 18, No. 21, Bureau of Educational Research Bulletin No. 5. Urbana: University of Illinois, 1921, p. 19.

[8]Odell, Charles W. "The use of intelligence tests as a basis of school organization and instruction." University of Illinois Bulletin Vol. 20, No. 17, Bureau of Educational Research Bulletin, No. 12. Urbana: University of Illinois, 1922. 78 p.

and one-half years of mental age. Investigation revealed that the teachers of these pupils had given instruction which incidentally functioned as coaching and increased the scores on the second trial of the test.

In an unpublished study made by Mr. H. N. Glick, deliberate coaching on the Army Group Intelligence Scale Alpha was given to a number of pupils. The increases in the scores when the test was repeated amounted in some cases to several hundred percent of the original scores.

In dealing with measures of achievement it is more difficult to demonstrate the presence of constant errors. Instruction is expected to result in an increase in achievement. However, the use of a standardized educational test implies the existence of standard conditions with respect to recency of instruction. Furthermore, in many cases we are measuring merely a sample of a pupil's achievement. In such cases our measurements are valid only if this sample is representative. Hence the increase in the scores yielded by a second application of an achievement test may represent a combination of true growth and spurious growth. If the instruction has been very recent in the field of the test or if it has been concentrated upon the particular sample of achievement which the test measures directly the increase in scores will represent, for the most part, spurious growth. Additional evidence on this point will be given in the next section.

When two dimensions of a pupil's ability are measured separately as in the case of both rate and comprehension of silent reading, we find that frequently the magnitude of one dimension is increased at the expense of the other. This may be due to the instruction which pupils have received or to other directions given them at the time of taking the test. Unless the two dimensions are interpreted together the effect of such compensating relation will be similar to that of a constant error.

4. **Evidence of constant errors in measures of progress in educational experimentation.** When we attempt to secure a measure of progress in achievement in a school subject by taking the difference between the averages (or medians) of two sets of scores, we frequently find evidence that one or both sets of scores involves an unknown constant error. Table II gives certain gains in achievement which were obtained in an experiment to determine the relative effect

TABLE II. TWO SETS OF GAINS IN ACHIEVEMENT WHICH INDICATE THE PRESENCE OF CONSTANT ERRORS IN CERTAIN SETS OF SCORES, FIFTH GRADE

Group	No. of Pupils	Reading Rate		Reading Comprehension		Arithmetic	
		I	II	I	II	I	II
I	70	27.93	−15.78	.96	.35	23.82	21.45
II	72	3.67	22.11	1.21	1.86	14.72	5.44
III	326	−4.77	33.25	.92	2.06	12.07	6.36
IV	133	−6.60	22.90	.82	.95	17.04	10.09
V	157	9.29	27.35	1.48	2.12	10.65	5.83
VI	143	−9.26	41.48	.08	2.36	4.69	5.38

of the number of sections into which a class was divided.[9] The six experimental groups were taught under the same conditions with the one exception regarding the number of sections into which the classes were divided. The tests used were Monroe's Standardized Silent Reading Test I, Revised, and Monroe's General Survey Scale in Arithmetic. Form 1 of these tests was given early in October, Form 2, the first of February, and Form 1 was repeated early the following May. The first gains were calculated by subtracting the average of the October scores from that of the February scores; the second, by subtracting the average of the February scores from that of the May scores. The two forms of these tests have been shown to be slightly lacking in equivalence, especially in the case of reading rate.[10] The gains in Table II, however, are evidence of the presence of constant errors in addition to those resulting from the slight non-equivalence of the different forms.

On the basis of our knowledge of the effect of practise we should expect the first gains to be larger than the second gains unless the variations of experimental conditions materially influenced the achievements of the pupils, which is extremely unlikely. We find in both reading rate and reading comprehension that the first gains are frequently less than the second. In three cases the first gain for reading rate is negative. In arithmetic the first gains are larger

[9]Monroe, Walter S. "Relation of sectioning a class to the effectiveness of instruction." University of Illinois Bulletin, Vol. 20, No. 11. Bureau of Educational Research Bulletin, No. 11. Urbana: University of Illinois, 1922. 18p.

[10]Monroe, Walter S. "The Illinois Examination," University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 12-18.

than the second in all cases except one. The smaller gains during the first semester than the second and particularly the negative gains are evidence of the presence of a constant error in at least one of the sets of scores from which the gains were computed. The gain in reading rate shown for Group I is interesting; from October to February there is a very marked increase in rate; for the second semester the gain is negative. This suggests that the average February score was too large, i.e., it involved a positive constant error.

Similar evidence of the presence of constant errors in measures of achievement is found in a recent study of the relation of class size to school efficiency.[11] In this investigation, as in the one just described, experimental groups were arranged in pairs with the experimental conditions alternating in the two semesters. Especially in Grades V and VII the relative magnitude of gains made in the different semesters indicates the presence of a constant error in at least one set of scores from which the gains were computed.

In another investigation[12] conducted by the Bureau of Educational Research the average increases in mental age during a period of six months for two groups of children, each numbering about 3000, were found to be .4 years and .9 years. During the next six months for the same two groups the increases were 1.4 years and 1.0 years respectively. The normal increase in mental age during either of these intervals is of course six months. The obtained increase for the first period might be expected to be somewhat greater because of the presence of a constant error introduced by the general practise effect. However, in one case the difference between the first and second trial scores is less than six months and in both the increase is less than the corresponding differences between the second and third trial scores. No explanation was found for these inconsistent gains but they are evidence that in some way an unknown constant error was introduced in some if not in all of the scores. The facts of this illustration become even more striking when we note that the total of the two gains for the first group is 1.8 years and that for the second 1.9. Thus, when the total interval of twelve

[11]"Relation of size of class to school efficiency." University of Illinois Bulletin, Vol. 19, No. 45, Bureau of Educational Research Bulletin No. 10. Urbana: University of Illinois, 1922, p. 20.

[12]Odell, Charles W. "The use of intelligence tests as a basis of school organization and instruction." University of Illinois Bulletin, Vol. 20, No. 17, Bureau of Educational Research Bulletin No. 12. Urbana: University of Illinois, 1922. 78 p.

months is considered, the total increase in mental age is approximately the same for the two groups. On the other hand, if the two intervals of six months are taken, the increases in the mental age are radically different for the two groups.

In the same investigation if only the scores yielded by the Illinois General Intelligence Scale are considered, the gain between the first and second testings is 1.1 years. For the second period it is 1.4 years. A constant error due to practise effect is expected in these gains but it is surprising to find that the second gain, which is the difference between the second and third trial scores, involves the larger error.

In each of these illustrations we have evidence of the presence of a constant error for which the cause is obscure. Furthermore, the exact magnitude of the constant error is unknown. The obscurity of the cause is due in part to the large number of teachers and pupils participating in each of these educational experiments. The constant errors may have been due to changes in the interest and attitude of the teachers and pupils toward the test. However, it was not possible to secure any direct evidence on this point. The fact that the cause is obscure makes the possible presence of constant errors in such data a serious matter and tends to arouse suspicions regarding the accuracy of measurements of ability in large cooperative experiments.

5. **Evidence of constant errors in subjective scoring.** The evidence cited in the preceding pages has related to testing conditions. The scoring of the tests used was highly objective. In case the scoring of the test papers is not objective it is necessary to consider also the constant errors which may be introduced in this process. In the marking of examination papers and other pupil performances where the scorer is asked to exercise judgment, much evidence has been collected to show that two persons differ widely in the scores which they assign to the same pupil performances. These differences are due in part to the presence of a constant error resulting from the fact that one of the scorers tends to be more liberal than the other. In a recent investigation[18] several sets of pupil performances for which the scoring was rather highly subjective, were scored inde-

[18]Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22. Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922. 52 p.

TABLE III. SUBJECTIVITY OF SCORING REPRODUCTIONS BY THE WORD-COUNTING METHOD

Test	Form	Grade	No. of scores	Scorers	Difference of average scores
Memory	I	IV	92	Y—C	− 9.9
Memory	I	IV	27	Y—K	− 5.1
Memory	II	IV	116	Y—C	− 2.0
Memory	I	VII	123	Y—K	− 7.5
Memory	II	VII	100	Y—C	− 8.2
Memory	II	VII	31	Y—K	+ 4.1
Reproduction	I	IV	94	L—K	+ 6.8
Reproduction	II	IV	31	L—C	− 1.6
Reproduction	II	IV	68	L—K	+ 4.7
Reproduction	I	VII	117	M—F	− 0.5
Reproduction	II	VII	113	F—C	− 6.0
Brown	I	IV	111	T—My	+12.8
Brown	II	IV	110	T—My	+ 6.9
Starch (No. 7)	I	VII	119	M—C	− 5.8
Starch (No. 6)	II	VII	121	M—C	− 2.0

pendently by two persons under the supervision of a third. A part of one table is reproduced from this report to furnish evidence of the presence of a constant error in the scores assigned by one or both of the scorers. The entries in the column headed "difference of average scores" were obtained by subtracting the average of the scores assigned by the second scorer from the average of those assigned by the first scorer. Some of these differences are relatively large. It appears that the scorer is not always consistent with respect to his constant error. Scorers Y and K show positive differences for one set of papers and negative differences for another set.

In the same investigation, eighty-six compositions were rated independently by two persons using the Willing Scale for Measuring Written Composition. The difference between the averages of their scores was 6.7.

Constant errors in first trial scores. As we have already indicated, first trial scores may involve constant errors. If there have been any departures from standard testing conditions we may expect to find the scores yielded too high or too low. It is possible to coach pupils for the first administration of a test as well as for later administrations. It may happen that where there has been no intentional coaching the instruction which they have received immediately prior to the taking of the test has served as preparation for the test.

Furthermore, if the norms are for pupils who are relatively unacquainted with testing procedure, test scores made by pupils who are accustomed to taking tests will involve a constant error with reference to these norms. At first the norms for our standardized educational tests were based upon scores obtained from pupils who had little or no experience in the taking of tests. This was necessarily so because such tests were new. As tests have become more widely used this factor of the testing conditions has changed, and it is probably true that norms for tests which have been recently standardized are based upon scores from many pupils who are familiar with general testing procedure. However, we have no specifications concerning the degree of acquaintance with the testing procedure for which the norms are stated.

In addition to the influence of instruction and acquaintance with testing procedures, constant errors may be introduced in first trial scores by the attitude of the pupils toward the test, by the way in which the test is explained to the pupils, and by a number of other factors which are subject to only partial control. In the case of handwriting the performances of pupils are very easily influenced by the type of directions given them. For example, in response to the instructions "Write as fast as you can" one college sophomore increased her rate of writing 77 letters per minute over her rate when writing for highest quality. One investigator[14] has presented evidence which shows that if pupils know they are being tested they will tend to write much more slowly than their normal rate. This reduction in rate is usually accompanied by an increase in quality. Similar results have been found for tests in other fields. The fact that test scores are influenced in this way by the directions given to pupils does not mean that they necessarily involve a constant error. It is only when these directions constitute departures from the standard testing conditions that we may expect constant errors. The evidence presented here merely shows what may happen when there are even slight departures from standard testing conditions.

Exact magnitude of constant errors can not be determined. In none of the cases cited to illustrate the presence of constant errors, was it possible to determine the exact magnitude of the constant error unless some basis for comparison was assumed. When a test

[14]Sackett, L. W. "Comparable measures of handwriting." School and Society, 4:640-45, October 21, 1916.

is repeated after a short interval of time the difference between the averages of the scores obtained from the two trials becomes the magnitude of the constant error in the second trial scores only if the first trial scores involve no constant error. Such an assumption may be justified in certain cases but one can never be certain that standard testing conditions prevailed in all details. Even when the examiner has exercised special care some of the more subtle factors of the testing conditions may not have been completely controlled. Unless good evidence can be produced in support of the assumption that the first trial scores involved a negligible constant error it is not safe to consider the difference between the averages of the two sets of scores as equivalent to the constant error. In more complex situations where a test is given three or more times for the purpose of measuring progress for two or more periods, it becomes more obvious that the exact magnitude of the constant error can not be determined. This condition has been indicated already in the evidence presented to show that constant errors were introduced in the data gathered in large cooperative educational experiments.

Altho one can not determine the exact magnitude of the constant error of measurement in a given case he can frequently present evidence to show that it probably does not exceed a certain amount. If his use of the data does not involve precise comparisons it may be possible to show that the constant error may be safely neglected. However, when precise comparisons are required and conclusions depend upon small differences between average or median scores the possible presence of constant errors makes such conclusions of doubtful validity.[15]

---

[15]In order to contrast the effects of the two types of errors upon the average and other derived measures, the consideration of the fourth question stated on page 7 with reference to constant errors is left until after the treatment of variable errors. The effect of both types of errors upon derived measures is considered in Chapter IV.

CHAPTER III

CAUSES, NATURE, AND MAGNITUDE OF VARIABLE ERRORS OF MEASUREMENT

1. Evidence of variable errors of measurement secured when a test is repeated. In order to secure evidence of the presence of variable errors of measurement it is necessary only to repeat a test after a short interval of time and compare the two scores of individual pupils. When this is done it is found that some pupils make a higher score on the first test and others on the second. In Table IV, two sets of scores yielded by the Monroe General Survey Scale in Arithmetic are given. The first pupil made a score of 51 on the first trial and 59 on the second. The difference in the two scores is — 8. Most of the differences are small. A few are relatively large. Approximately half are positive. The facts shown in this table are typical of the scores yielded by educational tests. For a few tests the scores involve smaller variable errors of measurement but for a number they are larger than in this illustration.

It should be noted that the differences between the two scores given in Table IV are not the variable errors of measurement. They are merely indicative of the presence of such errors and, in a crude way, of their magnitude. Neither set of scores can be considered true scores. Both are subject to variable errors and also, possibly, to an unknown constant error. In order to obtain the exact magnitude of the variable error of measurement for a given pupil it would be necessary to secure a true score and to subtract the obtained score from it. Such differences, when corrected for the constant error, would be the variable errors of measurement.

Method of describing the variable errors of measurement. As we have already indicated, it is impossible to determine the pupil's true score (see page 9). Furthermore, we always find variable errors and constant errors occuring in combination. It is, however, possible to secure a description of the magnitude of the variable errors of measurement which may be expected in the scores yielded by a given educational test when it is administered under standard testing conditions. Two sets of scores such as given in Table IV furnish the data upon which this description is based. These are obtained by

TABLE IV. SCORES YIELDED BY TWO APPLICATIONS OF MONROE'S GENERAL SURVEY SCALE IN ARITHMETIC TO A GROUP OF FIFTH GRADE PUPILS

Form I	Form II	Difference	Form I	Form II	Difference
51	59	− 8	46	58	−12
49	60	−11	49	54	− 5
46	60	−14	60	71	−11
77	84	− 7	43	41	+ 2
42	43	− 1	45	40	+ 5
43	43	0	30	24	+ 6
51	63	−12	28	23	+ 5
33	36	− 3	46	38	+ 8
41	48	− 7	34	32	+ 2
40	46	− 6	59	56	+ 3
39	53	−14	63	72	− 9
35	47	−12	42	45	− 3
42	47	− 5	45	53	− 8
21	25	− 4	48	73	−25
113	109	+ 4	51	59	− 8
42	35	+ 7	53	56	− 3
11	8	+ 3	68	68	0
23	21	+ 2	24	19	+ 5
45	38	+ 7	45	66	−21
54	49	+ 5	21	21	0
21	31	−10	37	36	+ 1
53	48	+ 5	38	39	− 1
43	53	−10	39	50	−11
106	86	+20	50	64	−14
27	19	+ 8	30	43	−13
46	45	+ 1	33	42	− 9
42	29	+13	65	74	− 9
45	62	−17	69	86	−17
55	54	+ 1	51	59	− 8
46	35	+11	43	43	0
38	32	+ 6	53	52	+ 1
17	15	+ 2	38	35	+ 3
53	61	− 8	62	72	−10
52	65	−13	77	75	+ 2
50	58	− 8	85	76	+ 9
41	43	− 2	111	95	+16
37	48	−11	70	66	+ 4
26	34	− 8	107	84	+23
57	65	− 8	9	8	+ 1
51	59	− 8	27	25	+ 2
34	41	− 7	39	32	+ 7
26	36	−10	25	23	+ 2
22	24	− 2	59	55	+ 4
59	64	− 5	57	47	+10
49	37	+12	41	36	+ 5
61	75	−14	56	62	− 6

two applications of the same test or of duplicate forms of a test to a group of representative pupils. These two applications should be separated by only a small interval of time. One type of description of the magnitude of the variable error of measurement is obtained by calculating the coefficient of correlation between these two sets of scores. This, when applied to an educational test, is called the coefficient of reliability. It indicates in a rough way the magnitude of the variable error of measurement, and is unaffected by the presence of any constant error of measurement in the two sets of scores from which it is calculated.

The coefficient of reliability an unsatisfactory description of variable errors of measurement. Altho the coefficient of reliability is an index of the variable error of measurement, a given coefficient, say .85, can not be interpreted directly in terms of the magnitude of these variable errors of measurement. Experienced persons are able to attach a reasonably concrete meaning to a given reliability coefficient but to an inexperienced person a coefficient of reliability, say .72 or .95, can have little more than a very general meaning.

Under certain conditions reliability coefficients furnish us with an index of the relative magnitude of the variable errors to be expected in the scores yielded by different tests. For example, if the reliability coefficient for one test is .65 and for another, .85, we should expect to find the variable errors of measurement for the second test much smaller. However, considerable caution must be exercised in comparing coefficients of reliability. The writer has shown[1] that the correlation of the two scores yielded by a given test is much smaller when the scores are taken from a single grade than when taken from a sequence of two or more grades. In one illustration when the scores were assembled separately for half-grade groups the highest coefficient of correlation was .57. For one half grade it was .12 and for another .27. When the scores for all grades from III-B to VIII-A, inclusive, were assembled the coefficient of correlation was .76. In this illustration there were from two hundred to four hundred cases in each half grade. Hence the variations can not be explained on the basis of sampling.

Probable error of measurement used to describe the variable error. Altho the coefficient of reliability is unsatisfactory it may

[1]Monroe, Walter S. An Introduction to the Theory of Educational Measurements. Boston: Houghton Mifflin Company, 1923, p. 356.

be used as a basis for calculating the probable error of measurement. The formula[2] for this is given below.

$$P.E._M = .6745 \frac{\sigma_1 + \sigma_2}{2} \sqrt{1 - r_{12}}$$

In this formula r_{12} is the coefficient of correlation between first and second trial scores. σ_1 is the standard deviation of the distribution of the first trial scores and σ_2 a corresponding measure for the second trial scores.

It should be noted that the probable error of measurement does not give the magnitude of the variable error of measurement for any one pupil. It gives merely the limits between which we may expect to find 50 percent of the variable errors of measurement of a given group of scores. For example, the probable error of measurement for the rate score yielded by the Courtis Silent Reading Test No. 2 has been found to be 19.3.[3] This means that 50 percent of the variable errors of measurement were greater than 19.3, approximately one-half of them being positive. It also means that 50 percent of them were not larger than 19.3 nor smaller than — 19.3. In the case of a given pupil we can state only the chances that the probable variable error of measurement of his score does not exceed certain limits; as, for example, in the Courtis Silent Reading Test referred to, the chances were just even that the variable error of measurement in his score was not larger than ± 19.3. The chances are 4.6 to 1 that the variable error of measurement of his score is between — 38.6 and +38.6. The chances for other limits also may be stated.

The magnitude of variable errors to be expected in educational measurements. In the data given to show the presence of both constant and variable errors in our educational measurements their magnitude has been indicated. It is clear that they are much greater than the corresponding errors in physical measurements. In another place the writer has discussed the relative magnitude of the errors in the scores yielded by standardized tests and the errors in the

---

[2]For an explanation of this formula see, Monroe, Walter S. An Introduction to the Theory of Educational Measurements. Boston: Houghton Mifflin Company, 1923, p. 347-56.

[3]Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p. 33.

grades assigned to examination papers.[4] The evidence presented indicates that the variable errors of measurement for a number of widely used standardized educational tests are only slightly less than the variable errors of measurement for written examinations. Some additional data with reference to the variable errors of the scores yielded by standardized tests will be helpful in arriving at the true understanding of their magnitude.

In a critical study of a group of silent reading tests[5] it was shown that the probable error of measurement for some tests was greater than 25 percent of the average score. In fact, for Brown's Silent Reading Test it was found to be more than 50 percent. In the tests which make up the Illinois Examination, only twelve of forty-two probable errors of measurement which were calculated were greater than 10 percent of the average score.[6] The authors of the Stanford Achievement Test[7] announce that the probable error of measurement for this battery of tests is approximately two months of educational achievement. The coefficients of reliability are high. It is likely that these authors have succeeded in reducing the variable errors of measurement to a lower minimum than has been secured by others. This has been accomplished in part through extending the length of the test.

Using scores which were the medians of eight independent ratings of English compositions by means of the Nassau County Supplement to the Hillegas Scale, Hudelson[8] has given coefficients of reliability ranging from .69 to .84. The writer has estimated that if the ratings of a single judge had been used the coefficients of correlation would have been in the neighborhood of .40 instead of ranging

---

[4]Monroe, Walter S., and Souders, L. B. "The present status of written examinations." University of Illinois Bulletin, Vol. XXI, No. 13. Bureau of Educational Research Bulletin No. 17. Urbana: University of Illinois.

[5]Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p. 52.

[6]Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 49.

[7]Kelley, T. L., Ruch, G. M., and Terman, L. M. Stanford Achievement Test Manual of Directions. Yonkers: World Book Company, 1923.

[8]Hudelson, Earl. "English composition, its aims, methods, and measurement." Twenty-second Yearbook of the National Society for the Study of Education, Part I. Bloomington, Illinois: Public School Publishing Company, 1923, p. 62.

from .69 to .84, and the probable error of measurement would have been a little more than six-tenths of one step of the scale used. This may appear relatively small but when we examine the norms we find that the unit used is relatively large. The average increase in norms from the fourth to twelfth grades, inclusive, is only slightly more than four-tenths of a unit per year. Between the eighth and ninth grades the increase is only two-tenths of a unit. The greatest yearly increase is six-tenths of a unit. Thus we have here a probable variable error of measurement which is relatively large.

CHAPTER IV

THE EFFECT OF CONSTANT AND VARIABLE ERRORS UPON DERIVED MEASURES

The effect of errors of measurement upon derived measures. There seems to be a prevailing idea that the effect of errors of measurement upon such derived measures as the average, median, standard deviation and coefficient of correlation, may be safely neglected if the derived measure is based upon a sufficiently large number of cases. This is only partially true. A constant error in the original data makes the average in error by the amount of the constant error. Any increase in the number of cases has no effect upon the magnitude of the constant error. It can not be eliminated or even reduced unless we are able to determine its magnitude, in which case it may be subtracted from the average. The same situation prevails for the median. However, a constant error does not affect the standard deviation and other measures of variability. Neither does it affect the coefficient of correlation.

As we have already pointed out in the preceding pages, we are seldom able to determine even approximately the magnitude of the constant errors of the scores yielded by educational tests. We have evidence only of their presence. Hence, it is impossible to make any accurate correction for a constant error. It has been estimated that second trial scores are about 10 percent larger than first trial scores but studies of different tests indicate that this constant error of measurement varies widely. For some tests the difference between first trial scores and second trial scores is much less than for others. It is also doubtless less for some groups of pupils than for others. When pupils are acquainted with the testing procedure the increase of second trial scores over first trial scores may be very slight, especially if the children are acquainted with other tests having similar structure. When compared with first trial scores, third trial scores involve a somewhat larger constant error than those obtained from the second trial, and beyond the third trial it is likely that there is some increase.

Unlike constant errors of measurement variable errors tend to neutralize each other in the average. The reason for this is easily understood because approximately one-half of the variable errors of

measurement are negative and the other half positive. If we increase the number of cases the magniture of the variable errors of measurement in the average is decreased. The relation is given by the following formula:

$$\text{P.E.}_{M}\ \text{average} = \frac{\text{P.E.}_{M}}{\sqrt{N}}$$

It should be noted that the error of the average due to the presence of variable errors of measurement in the data can not be explicitly defined. It is necessary to describe it in terms of the probable error (P.E._M average). The above formula gives the limits between which the chances are even that the error of the average will fall.

Variable errors of measurement tend to make the standard deviation and other measures of variability larger than they would be otherwise. The relation between the obtained standard deviation and the true standard deviation is given by the following formula:

$$\sigma_{\text{true}} = \sigma_{\text{obtained}}\sqrt{r_{12}}$$

In this formula r_{12} is the coefficient of reliability of the scores concerned. Since N does not appear in this formula it follows that increasing the number of cases does not have any effect upon the variable error of measurement of a measure of variability.

The presence of variable errors of measurement in our data always tends to decrease the coefficient of correlation.[1] If each of the two sets of facts whose relationship we are studying has been measured in duplicate, it is possible to correct for the effect of these variable errors of measurement. For example, if it is desired to secure the true correlation between ability to reproduce a selection read and ability to answer questions upon it, we may secure a corrected coefficient of correlation by measuring each of these abilities twice. One formula which has been used for this purpose is the following:

$$r_{pq} = \frac{\sqrt{(r_{p_1q_2})\ (r_{p_2q_1})}}{\sqrt{(r_{p_1p_2})\ (r_{q_1q_2})}}$$

r_{pq} here indicates the true correlation between two series of measures, p and q, of the facts A and B.

p_1 and p_2 are two independent measures of A.

q_1 and q_2 are two independent measures of B.

[1]Thorndike, E. L. Theory of Social Measurements. New York: Teachers College, Columbia University, 1916, second edition, p. 178.

$r_{p_1q_2}$ is the correlation obtained from the first measure of A and the second measure of B.

$r_{p_2q_1}$ is the correlation obtained from the second measure of A and the first measure of B.

$r_{p_1p_2}$ is the correlation between the two measures of A.

$r_{q_1q_2}$ is the correlation between the two measures of B.

In a recent study[2] it was desired to secure the correlation between the comprehension in silent reading as measured by Monroe's Standardized Silent Reading Tests with the scores yielded by a test of memory. It was recognized that both sets of scores involved variable errors of measurement which would materially decrease the magnitude of the coefficient of correlation. For this reason it was arranged to measure each trait in duplicate. The coefficients of correlation obtained from correlating each measure of comprehension yielded by a silent reading test with the two measures of memory were .31, .33, .31, and .35. The coefficient of reliability for the memory test was .35 and for the silent reading test, .73. In this illustration $r_{p_1p_2}$ equals .65, $r_{q_1q_2}$ equals .35, $r_{p_1q_2}$ equals .33, and $r_{p_2q_1}$ equals .31. Substituting these values in the formula given above we obtain for the corrected coefficient of correlation between comprehension and memory a value of .67. This gives some indication of the effect of the variable errors of measurement in these two sets of scores upon the coefficient of correlation between comprehension and memory.

---

[2]Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin, No. 8. Urbana: University of Illinois, 1922, p. 41.

CHAPTER V

EFFECT OF ERRORS UPON USE OF EDUCATIONAL MEASUREMENTS

The attitude toward educational measurements as affected by the recognition of errors. The effective use of any instrument depends upon a frank recognition of its limitations. Standardized educational tests are no exception. They have been advertised by their authors and by others as measuring instruments which are distinctly superior to ordinary written examinations. At first attention was centered upon their use and few studies were made of the errors involved in the scores yielded. It has been apparent for some time, however, that test scores are subject to errors which sometimes are astonishingly large. This bulletin has been written to call attention to the nature of these errors, their magnitude, and their effect upon the average and other derived measures.

Those of us who have not been concerned with the errors involved in educational measurements may tend to feel doubtful of the value of standardized tests after realizing the nature and magnitude of the errors encountered. In the judgment of the writer, this effect should not be produced. We cannot expect to make our use of educational tests most effective until we are informed concerning the limitations of the measures yielded. A frank recognition of the presence of both constant and variable errors should enable the users of educational tests to do their work more efficiently. In many cases they can avoid erroneous interpretations which they otherwise would make. Furthermore, it is only by understanding the nature of the errors which are likely to be encountered that we can take steps to reduce them to the lowest possible minimum, and be aided in the construction of improved measuring instruments because of our knowledge of the defects of the present ones.

The writer does not advise the discontinuance of the use of standardized educational tests because the measures yielded have been shown to involve both variable and constant errors larger than many of us supposed. There is abundant evidence to show that the use of educational tests in our schools is increasing their efficiency. Still greater improvement may be expected when measuring instruments are used more intelligently.

BULLETIN NO. 16

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

AN ANNOTATED BIBLIOGRAPHY DEALING WITH THE CLASSIFICATION AND INSTRUCTION OF PUPILS TO PROVIDE FOR INDIVIDUAL DIFFERENCES

by

CHARLES W. ODELL
Associate, Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA
1923

PREFACE

The making of adequate provisions for individual differences represents one of the most important problems which educators are now facing. The need is beginning to be generally recognized and numerous procedures are being tried out in our schools. Many reports of attempts to provide for individual differences are being made. In order that one who is interested in making such provisions may profit by the experience of others it is necessary that he acquaint himself with the work which they have done. In order to assist the practical school man in this endeavor Dr. C. W. Odell has brought together an extended bibliography on "The Classification and Instruction of Pupils to Provide for Individual Differences." To each of the references he has added a brief annotation which indicates the character of the information which will be found in the reference. Altho the list probably is not complete it is believed to be more so than any other bibliography on this topic which is now available.

WALTER S. MONROE, *Director.*

November 10, 1923

An Annotated Bibliography Dealing with the Classification and Instruction of Pupils to Provide for Individual Differences

The bibliography which follows was prepared by the writer in connection with his work described in a previous bulletin of the Bureau of Educational Research entitled "The Use of Intelligence Tests as a Basis of School Organization and Instruction," and may be considered in the light of an appendix to that bulletin. It has, however, been revised and brought up to date, that is practically to the end of the school year of 1922-23. No claim is made for its absolute completeness nor that the references given are more important than those not included. An effort has been made to list all important and most unimportant references that have appeared in the twenty-five or thirty leading educational periodicals of this country. In addition to these references many others are included from books, school reports, bulletins, proceedings of various educational bodies, etc. The writer did not make a thoro search through all of the reports of city superintendents and other school officials, proceedings of state and other teachers' associations and various other similar sources.

The references given have been classified into two main divisions. The first part of the bibliography deals with discussions of and provisions for individual differences that involve the use of standardized tests of general intelligence or achievement. The second part contains references about provisions and discussions that do not involve the use of such tests. Articles that deal with both sorts of classification, that based on the results of tests and that upon some other basis, are included in the first part.

PART I.

1. Adler, Martha. "Mental tests used as a basis for the classification of school children," Journal of Educational Psychology, 5:22-28, January, 1914.

In a New York school 160 first and fourth-grade pupils were tested with the Goddard Revision. Almost 50 percent of those tested were placed in advanced sections and of these about 60 percent gained one-half year's time. Only two were retarded.

2. ALEXANDER, J. M. "Binet-Simon test in practical use in the public schools of Hinsdale, Illinois," Elementary School Journal, 21: 146-48, October, 1920.

Fifty-three first-grade pupils were tested and rearranged on the basis of their ability in three sections, the best of which did extra work. Also special classes in other grades were tested and handled in accordance with the results.

3. ALMACK, J. C. and J. L. "Administrative problems connected with gifted children," Educational Administration and Supervision, 8:129-36, March, 1922.

In the junior high school of Eugene, Oregon, about 800 pupils were tested with the Army Alpha, Otis and Stanford Revision. Fifty-one with I. Q.'s above 100 were selected for superior sections.

4. ANDERSON, ROSE G. "Methods and results of mental surveys," Journal of Applied Psychology, 6:1-28, March, 1922.

A general discussion of a number of mental surveys and of the number of pupils selected as feeble-minded. This varied from .16 percent to 6.4 percent.

5. ARMENTROUT, W. D. "Classification of junior high-school pupils by the Otis scale," Journal of Educational Psychology, 11:165-68, March, 1920.

Four hundred junior high-school pupils in Lawrence, Kansas, were tested with the Otis scale and three or four groups in each half year selected according to their I.Q.'s.

6. ARMENTROUT, W. D. "Grouping pupils by intelligence tests," School Review, 28:249-51, April, 1920.

Another account of procedure in the Lawrence, Kansas, Junior High School.

7. ARMENTROUT, W. D. "Classification and promotion of pupils," Education, 42:506-12, April, 1922.

A discussion of several flexible plans. Advocates seven, eight, and nine year courses for the elementary school with classification based upon intelligence, physical and educational tests, school marks and teachers' estimates.

8. ARMENTROUT, W. D. "Classification of junior high-school pupils by the Otis scale," Education, 43:83-87, October, 1922.

Another account of the classification of the junior high-school pupils at Lawrence, Kansas.

9. ARTHUR, GRACE. "An application of intelligence tests to the problem of school retardation," School and Society, 10:614-20, November 22, 1919.

In a St. Paul elementary school, group and individual tests were used and pupils placed according to the results. In two years failures were reduced from 11 to 2.9 percent, special promotions increased from 1.5 to 9.4 percent and $4000 saved.

10. AYER, F. C. "The present status of promotional plans in city schools," American School Board Journal, 66:37-39, April, 1923.

Gives questionnaire replies from 124 cities. Thirty-six different plans are listed. The cities average about ten plans apiece. It was found that there is little evidence as to the value of the various plans.

11. BADANES, JULIE E. "The first practical steps in selecting gifted children in a large city school." New York: Continental Printing Company, 1921. 22p.

A brief history of intelligence testing discussing especially the work and writings of Burt, Meumann, Allen, Pearson and others.

12. BAER, J. A. "Individual differences among pupils," Cleveland, Ohio; Cuyahoga County Public Schools, 1922. 28p.

A general discussion of the individual differences that exist followed by suggestions as to different bases and plans of classification and what the teacher can do to look after individual differences in the different subjects. This is followed by a short bibliography.

13. BAGLEY, W. C. "Educational determinism; or democracy and the I. Q.," School and Society, 15:373-84, April 8, 1922. Educational Administration and Supervision, 8:257-72, May, 1922.

A strong plea against classifying or instructing pupils according to their I.Q.'s as yielded by mental tests.

14. BAGLEY, W. C. "Professor Terman's determinism: A rejoinder," Journal of Educational Research, 6:371-85, December, 1922.

An answer to Terman's criticism of Bagley's address on Educational Determinism.

15. BAGLEY, W. C. "Educational determinism again: A rejoinder to Professor Whipple's reply," School and Society, 16:141-44, August 5, 1922.

A further argument against the use of intelligence test results for classifying school children.

16. BARTON, J. W. "School organization on an objective basis," Educational Administration and Supervision, 6:187-97, April, 1920.

In Elk River, Minnesota, 323 pupils in grades one to twelve were tested with the Kansas Silent Reading and an opposites test. Thirteen percent were given extra promotion on the basis of test results and of these all but three made good. Not a single one was injured physically.

17. Bates, Grace M. "The work of the students' welfare committee of Erasmus Hall High School," Bulletin of High Points in the Work of the High Schools of New York City, 5:20-24, June, 1923.

In this school, group intelligence test results are used to classify the pupils on the first day of actual high-school work. Whatever adjustments seem advisable are made later but it has been found that the original classification is fairly accurate.

18. Batson, W. H. "The South Dakota group intelligence test for high schools," School and Society, 15:311-15, March 18, 1922.

An account of testing almost 1500 pupils in South Dakota. Shows how test results compare with teachers' estimates, also the amount of variation between pupils and schools.

19. Berry, C. S. "Classification by tests of intelligence of ten thousand first-grade pupils," Journal of Educational Research, 6:185-203, October, 1922.

An account of the plan used in Detroit. First-grade pupils are tested and divided into X, Y and Z groups according to their ratings. The course of study differs for the three groups.

20. Biddle, Anne E. "The superior group," News Letter, 17:3-4, May, 1923.

A brief account of grouping by ability in the high schools of a certain city during the last several years with individual accounts of a number of bright pupils. Superior pupils were selected by the judgment of the teachers verified by intelligence tests. The writer is heartily in favor of the plan.

21. Bliss, D. C. "The application of standard measurements to school administration," Fifteenth Yearbook of the National Society for the Study of Education, Part I. Bloomington, Illinois: Public School Publishing Company, 1916, p. 69-77.

A general discussion of the question with some illustrations drawn from Montclair, New Jersey.

22. Book, W. F. "Variations in mental ability and its distribution among the school population of an Indiana county," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 100-31.

An account of the use of Indiana Schedule B in all the schools of a county. Discusses the differences found between pupils, the distribution of ability, the reliability of the scale used and the adjustments that should be made in the schools.

23. Bracewell, R. H. "The Freeman-Rugg general intelligence tests as an aid to economy in school administration," School Review, 29:460-66, June, 1921.

In the Burlington, Iowa, High School, pupils were tested with these tests and the average of the test results and teachers' rankings used for purposes of placement.

24. Branson, E. P. "An experiment in arranging high-school sections on the basis of general ability," Journal of Educational Research, 3:53-55, January, 1921.

In the Long Beach, California, High School, freshmen were grouped in English according to their scores on the Otis scale. This scale appeared to select inferior pupils better than superior ones, altho it was fairly efficient in both respects.

25. Breed, F. S. "Shall we classify pupils by intelligence tests?" School and Society, 15:406-9, April 15, 1922.

A general discussion listing a number of principles, cautions, conditions, and steps in using intelligence tests. The conclusion is that altho intelligence test results, the will, and the emotions, etc., are important factors in classifying, the achievement secured must be the final determining factor.

26. Breed, F. S. and Breslich, E. R. "Intelligence tests and the classification of pupils. II." School Review, 30:210-26, March, 1922.

An account of an experiment in the University of Chicago High School in which it was found that Otis test results divided pupils better than measures of their ability in arithmetic. The conclusion reached was that no test gives a reliable basis for permanent classification.

27. Brooks, S. S. "Some uses for intelligence tests," Journal of Educational Research, 5:217-38, March, 1922. Improving Schools by Standardized Tests. Boston: Houghton Mifflin Company, 1922. p. 96-123.

Recommends that pupils be placed by intelligence test results rather than by those from achievement tests.

28. Buckingham, B. R. "Suggestions for procedure following a testing program-I. Reclassification," Journal of Educational Research, 2:787-801, December, 1920.

States that intelligence tests alone are not a safe guide for promotion but should be supplemented by subject-matter tests and measures of non-intellectual qualities.

29. Burkard, W. E. "Grouping by ability in the Robert Treat Junior High, Newark," News Letter, 19:4-5, April, 1921.

In Newark there were accelerated, normal and retarded groups in grades I-IX based on measurements by a modified form of the Binet Scale. The same course of study was followed but at different rates. The plan appears to have given satisfaction.

30. Callihan, T. W. "An experiment in the use of intelligence tests as a basis for proper grouping and promotions in the eighth grade," Elementary School Journal, 21:465-69, February, 1921.

Almost 300 eighth-grade pupils were divided into sections according to ability. The basis was a combination of results from the Illinois General Intelligence Scale, Monroe's Silent Reading Test, and to some extent teachers' estimates and marks.

31. CAMPBELL, CORA. "Intelligence tests as a basis for classification," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 45-49.

As a result of a number of experiments Kansas City adopted the policy of classifying pupils upon the basis of individual intelligence tests given in the kindergarten and first grade, group tests given in other grades, and health, attendance, school attitude and certain other factors.

32. CAMPBELL, L. H. "Age gradation and grade grouping," American School Board Journal, 58:36, May, 1919.

States that in Providence, Rhode Island, pupils were grouped according to their age and placings verified by the Stanford Revision. Ten good results are listed.

33. CARBACK, CLARENCE. "Grouping of children by abilities and consequent change in school procedure. 3. Procedure in Philadelphia." Ninth Annual Schoolmen's Week Proceedings. University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 269-72.

States that in Philadelphia there are about twenty elementary schools with some or all pupils grouped according to their abilities. The bright pupils take an enriched curriculum.

34. CAYCO, F. and PRESSEY, S. L. "Three refinements of method in school surveys," Educational Administration and Supervision, 7:433-38, November, 1921.

A discussion of the need of relating age-grade, achievement and intelligence test data to each other. Also a brief account of testing the pupils of a small system.

35. CHASSELL, C. S. and CHASSELL, L. M. "A survey of the three first grades of the Horace Mann school by means of psychological tests and teachers' estimates and a statistical evaluation of the measures employed," Journal of Educational Psychology, 12:72-81, 243-52; February, May, 1921.

The average of results according to the Stanford Revision, Pressey and Myers Tests and teachers' estimates of maturity and reading ability was found and used as a basis of forming three fairly homogeneous groups.

36. CLERK, F. E. "Providing for individual differences by grouping by abilities. Organization and practical working of the plan." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. XXI, No. 37, Philadelphia, June 18, 1921, p. 243-49.

An account of procedure in Winchester, Virginia, in which teachers' judgments, school marks, achievement and intelligence tests, etc., are used in placing pupils. Accelerated, normal and slow groups are formed in high school as well as grades. A number of cautions are given.

37. CLEVELAND, ELIZABETH. "Detroit's experiment with gifted children," School and Society, 12:179-83, September 11, 1920.

The best five percent of seventh and eighth-grade pupils are selected and placed in special classes. The first basis of selection was the teacher's judgment, later the Binet Tests were used and finally group tests. The pupils selected gain a little time and take an enriched course.

38. CLEVELAND, ELIZABETH. "Some further studies of gifted children," Journal of Educational Research, 4:195-99, October, 1921.

A study of 144 bright and the same number of average children in Detroit. Health, home conditions, tastes, etc., were studied.

39. COLE, L. W. "Mental age and school entrance," School and Society, 8:418-19, October 5, 1918.

In Denver pupils entering the first grade are tested. It is found that the I.Q. of five-year-olds is high, of six-year-olds next and that of seven-year-olds lowest. The conclusion is that better results could be obtained by selecting new entrants by the Binet Scale than by chronological age.

40. COLE, L. W. "Prevention of the lockstep in schools," School and Society, 15:211-17, February 25, 1922.

An account of the use of the Binet, and Cole and Vincent Tests in the first grade. It is shown that six-year-old children are not a homogeneous group and that there is a great difference between the best and worst entering the first grade. States that classes should be organized on a mental-age basis.

41. COLVIN, S. S. "The present status of mental testing," Educational Review, 64:196-206, 320-37; October, November, 1922.

A brief account of the testing movement and present knowledge concerning it and a discussion of the use of tests for the classification of pupils, their validity, cautions to be observed in their use, etc.

42. COLVIN, S. S. et al. "Intelligence tests and their use," Twenty-first Yearbook of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company, 1922, 289 p.

The most complete discussion of the subject published. Part I deals with the nature, history and general principles of intelligence testing and Part II with the administrative use of intelligence tests. The latter part contains both theoretical discussion and actual accounts of what is being done.

43. COX, E. M. "Report of committee on promotions and rates of progress." Superintendent's Report, Oakland, California, 1917-18, p. 77-79.

History of acceleration, retardation, special progress, mental testing, special classes, etc., in Oakland. States that the results of special promotions, classes and mental testing have been good.

44. Cox, P. W. L. "Providing for individual differences by means of grouping by ability." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 233-44.

An account of grouping pupils in junior and senior high schools according to results from the Otis Test, various achievement tests and their interests and aptitudes.

45. Coxe, W. W. "School variation in general intelligence," Journal of Educational Research, 4:187-94, October, 1921.

Gives results of giving the Otis Test in twenty-four Cincinnati elementary schools as a basis for selecting pupils for a special six-year classical course. Great variation in intelligence was found.

46. Cummins, R. A. "Educational measuring sticks and their uses," American School Board Journal, 63:33-5, August, 1921.

Contains a discussion of the use of intelligence and subject-matter tests.

47. Dawson, C. D. "Classification of kindergarten children for first grade by means of the Binet scale," Journal of Educational Research, 6:412-22, December, 1922.

In Grand Rapids 64 percent of 2000 kindergarten children were classified in the first grade on the basis of test results. Three groups were formed. Eighteen percent had to be shifted from the original classification.

48. Deamer, Arthur. "An experiment in acceleration," Fargo, North Dakota; Board of Education, 1919, 32 p.

Pupils for fast sections in the upper grades were selected according to their school marks. When measured with standardized tests they showed that their achievement was satisfactory. Likewise their high-school records were good.

49. Dickson, V. E. "The relation of mental testing to school administration." The Normal Seminar Bulletin A, No. 1. Cheney, Washington, June, 1917.

A study of the differences of pupils according to sex, chronological and mental age, I.Q., school mark, and teacher's estimate. The study shows a high correlation between mental ages and school marks.

50. Dickson, V. E. "What first-grade children can do in school as related to what is shown by mental tests," Journal of Educational Research, 2:475-80, June, 1920.

In Oakland, California, first-grade pupils are classified in three groups on the basis of test results and school marks. Pupils with a mental age below six cannot do regular first-grade work and should be put in a special group.

51. Dickson, V. E. "The use of group mental tests in the guidance of eighth-grade and high-school pupils," Journal of Educational Research, 2:601-10, October, 1920.

On the basis of scores on part of the Otis Scale and school marks an accelerated class was formed which covered eighth-grade work in fifty-five days, entered high school immediately and made better marks than the average. Two similar groups are also discussed.

52. Dickson, V. E. "The use of mental tests in school administration," Monograph, Vol. IV. Berkeley, California: Board of Education, June, 1922. 44 p.

Advocates the use of test results in determining the grade placement of pupils. Shows that individual intelligence tests are rather highly reliable and group tests fairly so, also that the I.Q. is a fairly safe measure of prognosis.

53. Dickson, V. E. "The treatment of gifted children in Oakland and in Berkeley." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 26-28.

A discussion of the basis of segregation and what should be done after segregation. Argues that children should be treated differently according to their differences in ability. Also that gifted children should not merely gain time but should also have a broader and surer foundation.

54. Dickson, V. E. Mental Tests and the Classroom Teacher. Yonkers-on-Hudson: World Book Company, 1923. 231 p.

This is, as its name indicates, a general discussion of the use of mental tests in school. The author maintains that group tests are fairly reliable for predicting school success and that pupils should be classified at the time of entering school on the basis of their ability. The results of mental tests should be an important factor in this classification.

55. Doten, Willard. "Tests for ability grouping," American School Board Journal, 63:37-38, October, 1921.

The Montclair, New Jersey Junior High School pupils were grouped according to the results from the National and Haggerty Tests. The work for the groups was differentiated. Satisfactory results followed.

56. Edmondson, Margaret B. "A mental survey of first-grade school pupils," Pedagogical Seminary, 27:354-70, December, 1920.

A discussion of the mental ages, I.Q.'s, school marks, chronological ages, and teachers' estimates of some grade pupils in Eugene, Oregon. Shows the lack of homogeneity of the group and the relationship between the various data.

57. Ettinger, W. L. "Economy in school administration," School and Society, 14:409-15, November 12, 1921.

An argument that much of the waste in school is due to crude classification.

58. Ettinger, W. L. "Crude classification causes waste," School Life, 7:66, November, 1921.

A general statement of the need of careful classification.

59. Fordyce, Charles. "Intelligence tests and classifying children in the elementary school," Journal of Educational Research, 4:40-42, June, 1921.

An account of the use of the Haggerty Test on over 1000 pupils. Discusses relation of test results and teachers' estimates.

60. FRANK, C. D. "A ray of light let in by a study of first-term failures," Bulletin of High Points in the Work of the High Schools of New York City, 5:6-7, May, 1923.

An account of an experiment showing that elementary-school marks furnish a better basis of predicting high-school marks than do intelligence test results.

61. FREEMAN, F. N. "Bearing of the results of mental tests on the development of the child," Scientific Monthly, 12:558-70, June, 1921.

States that the traditional method brings together pupils of very different mental ability, while grouping by mental ability brings together pupils whose intellectual processes are not equally developed. Ideal grouping would bring together pupils alike in both.

62. FREEMAN, F. N. "Bases on which students can be classified effectively," School Review, 29:735-45, December, 1921.

States that classification should be based upon mental ability, chronological and physical age, social and intellectual stages of development. Both tests and teachers' judgments should be used.

63. FRASIER, G. W. "The measurement of intelligence as an aid to administration," Educational Administration and Supervision, 6:361-66, October, 1920.

An account of administering the Stanford Revision in Spokane and giving pupils special promotions and demotions according to the results. In general the placement of pupils was satisfactory.

64. GLASS, J. M. "Classification of pupils in ability groups," School Review, 28:495-508, September, 1920.

In a Rochester junior-high school pupils were placed by results from several intelligence tests. Most of those so placed did satisfactory work but a few had to be shifted.

65. GODDARD, H. H. "Two thousand children tested by the Binet measuring scale for intelligence," Addresses and Proceedings of the National Education Association, 49:870-78, 1911.

An account of using the Binet scale to find New York children three years or more below normal.

66. GOLDSTONE, G. A. "Differentiation of method in teaching reading to slow and bright pupils," Bulletin of High Points in the Work of the High Schools of New York City, 5:11-14, April, 1923.

An account of an experiment in grouping high-school pupils into bright, medium and slow groups according to the results of the Thorndike-McCall Reading Scale. After the groups were formed each was taught by a different method but with practically the same subject matter. The results were satisfactory.

67. GRAY, P. L. and MARSDEN, R. E. "An application of intelligence tests," Journal of Experimental Pedagogy and Training College Record, 6:33-38, March 5, 1921.

An account of testing a small group of pupils with the Stanford Revision. Compares the results with those from examinations, teachers' estimates and vocabulary tests.

68. GREENBERG, B. B. "Intelligence tests as a basis for reclassification," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 55-58.

An account of the reorganization of a New York school on the basis of results from individual and group intelligence tests, several achievement tests, sensory and motor tests, etc. Reclassification on the basis of these tests resulted in increased homogeneity in each grade.

69. HAGGERTY, M. E. "Recent developments in measuring human capacities," Journal of Educational Research, 3:241-53, April, 1921.

States that intelligence must be supplemented by other traits as a basis for estimating capacities.

70. HANSON, W. A. "Mental measurement and the special class system in New London, Connecticut," American School Board Journal, 63:37, 119, September, 1921.

Twenty-seven hundred children were tested with the Dearborn and Haggerty Tests. Special classes were formed and the results also used to guide teachers in their instruction. In some cases individual tests were given.

71. HARPER, MAY M. "Intelligence tests as a basis for homogeneous grouping," Elementary School Journal, 22:781-82, June, 1922.

In a junior high school in Xenia, Ohio, pupils were classified in three groups according to the Haggerty and Terman tests. Reclassifications were made at the end of every two months. About half of the pupils were changed at the end of two months and about one-fourth at the end of four months.

72. HENMON, V. A. C. "The measurement of intelligence," School and Society, 13:151-58, February 5, 1921.

An argument in favor of the cautious use of tests for purposes of classification. States that these tests should measure social and mechanical as well as abstract intelligence.

73. HENRY, MARY B. "Mental testing as an aid in guidance and classification of school children." Department of Research of the Santa Ana Public Schools, Bulletin No. 1. Santa Ana, California: Board of Education, 1919. 23 p.

A general account of the use of mental tests in the Santa Ana schools. Discusses their agreement with teachers' estimates, their use for the purpose mentioned above, etc. Pupils are grouped in three sections on the basis of test results and other information.

74. Herrig, Anna B. "Promotions in the practise school as determined by the use of standard tests and educational measurements," Educational Administration and Supervision, 7:217-25, April, 1921.

A theoretical discussion.

75. Hines, H. C. "What Los Angeles is doing with the results of testing," Journal of Educational Research, 5:45-47, January, 1922.

Gives various details connected with the use of both intelligence and achievement tests in Los Angeles.

76. Hines, H. C. "Measuring the intelligence of school pupils," American School Board Journal, 64:35-37, 135, April, 1922.

A rather good theoretical discussion giving criticisms on both sides of the question and listing what are considered the best tests for the different grades.

77. Holley, C. E. "Mental tests for school use." Bureau of Educational Research Bulletin, No. 4, University of Illinois Bulletin, Vol. 17, No. 28, March 8, 1920. University of Illinois, 91 p.

A study of the use of a number of mental tests with some theoretical discussion.

78. Holmes, H. W. "The general philosophy of grading and promotion with relation to intelligence testing," School and Society, 15:457-61, April 29, 1922.

An argument for an enriched course for superior pupils.

79. Hughes, W. H. "Provisions for individual differences in high-school organization and administration," Journal of Educational Research, 5:62-71, January, 1922.

A questionnaire study of plans actually used by high schools.

80. Hunter, F. N. "Report of the superintendent of schools for the year 1917-18." Oakland, California: Board of Education, 1919. 353 p.

Contains considerable discussion of the Otis and Simon-Binet Tests, flexible promotion, etc.

81. Irving, M. Louise. "Classification into ability groups in Santa Rosa," Journal of Educational Research, 6:362-64, November, 1922.

In the junior high school of Santa Rosa classification is based upon two or three intelligence tests and several achievement tests.

82. Jahrling, Robert. "Educating gifted children in Hamburg," Pedagogical Seminary, 30:35-39, March, 1923.

Gifted children are selected by a combination of tests, teachers' estimates and school marks. A special course is provided for them.

83. Johnson, O. J. "Teachers' estimates of qualities of gifted pupils as related to classroom activities," School and Society, 17:466-69, April 28, 1923.

Thirty-three teachers of 900 pupils, mostly in high school, answered a questionnaire of ten questions on the general topic given above. The teachers favored separating bright pupils from others and giving them special work.

84. Jordan, R. H. "An example of classification by group tests," Educational Administration and Supervision, 6:198-201, April, 1920.

In a Minnesota grade school the seventh and eighth-year pupils were tested with several mental tests. Those above the 75 percentile of the next higher class were considered for extra promotion, those below the 25 percentile of the next lower class for failure or demotion.

85. Keener, E. E. "The use of measurements in a small city school system," Journal of Educational Research, 3:201-06, March, 1921.

This article mentions a few instances of readjustment in school based upon test results. The averages given are from the Richmond, Indiana, schools.

86. Kelley, T. L. "Again: educational determinism," Journal of Educational Research, 8:10-19, June, 1923.

The writer states that he is in sympathy with Bagley's belief in democracy and Terman's insistence upon the reality of individual differences. His conclusion is that instruction should be properly differentiated.

87. Kent, R. A. "An experiment in the grading and placing of children," American School Board Journal, 62:40-41, February, 1921.

In Duluth, pupils were placed according to results from the Otis test. Discipline was made easier, interest was increased, better work was done and other good results secured.

88. Ketner, Sarah P. "Grouping by standardized tests for instructional purposes," Journal of Educational Research, 2:620-25, October, 1920.

A Denver school grouped pupils in reading, arithmetic and penmanship according to the results of tests. The experiment seemed to be successful and the plan was to be extended to other schools.

89. Kuhlmann, F. "What constitutes feeble-mindedness," Journal of Psycho-asthenics, 19:214-36, June, 1915.

Reaches the conclusion that the I.Q. is the most reliable criterion of an individual's grade of intelligence.

90. KUNTZ, E. E. "Grouping of children by abilities and consequent changes in school procedure. 3. Procedure in Lansford." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 267-69.

The Haggerty, National and Terman Tests were used as a basis for grouping pupils. Two or three sections were formed in each of the lower grades and more time given to slow pupils in the upper grades.

91. KYTE, G. C. "An experiment in the education of gifted children in the first grade," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 71-80.

In a public school at Berkeley superior first-grade children are selected on the basis of results from the Stanford Revision and certain other data and allowed to gain time. A reading test showed that nothing was lost in achievement. On the whole the plan was satisfactory.

92. LAWS, A. R. and BOWIE, S. "Intelligence tests in examinations for junior scholarships," Journal of Experimental Pedagogy and Training College Record, 6:155-69, December 5, 1921.

Found that the correlation between intelligence test results and terminal marks ranged from —.04 to +.96. Concludes that no one test is sufficient as a basis for judging pupils.

93. LAYTON, W. K. "The intelligence testing program of the Detroit public schools," School and Society, 15:368-72, April 1, 1922.

Much the same as contained in the reference by same author in the Twenty-first Yearbook. Especially describes the X, Y and Z groups in the first grade, membership in which is determined by test results and which have different work.

94. LINDSAY, M. D. and GAMSBY, R. S. "Where test scores and teachers' marks disagree," School Review, 29:679-87, November, 1921.

An account of the use of the Terman Test in Palo Alto Union High School. It was found that in all cases of poor work by superior students the teachers could have secured better work if they had known the pupils' ability.

95. LLOYD, S. M. and ULLRICH, O. A., JR. "The progress of pupils in an ungraded class," Psychological Clinic, 11:276-87, February 15, 1918.

Several subject-matter tests were used with the pupils of an ungraded room in Austin. Pupils gained by being placed in this room.

96. LOWELL, FRANCES. "An experiment in classifying primary grade children by mental age," Journal of Applied Psychology, 6:276-90, September, 1922.

In the practise school of the City Normal School of Rochester, pupils of grades I to III were grouped with the Kuhlmann Revision as a chief basis.

97. Madsen, I. N. "Interpreting achievement in school in terms of intelligence," School and Society, 14:59-60, July 30, 1921.

A rather short discussion of the E. Q., A. Q. and I. Q.

98. Madsen, I. N. "Intelligence as a factor in school progress," School and Society, 15:283-88, March 11, 1922. American School Board Journal, 64:37-38, June, 1922.

A number of towns, mostly in Idaho, gave the Haggerty Test in the elementary school and the Army Alpha in high school to over 12,000 pupils. It was found that the I.Q.'s of the same age group varied directly with the grade placement.

99. Marshall, Jessica. "Using the results of testing," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 49-55.

An account of the use of tests for purposes of placing pupils in a Toledo school. In the lower grades classification was chiefly according to the results of various intelligence tests, in the upper grades according to those from achievement tests.

100. Martens, Elise H. "Better classification," The Journal of the National Education Association, 12:174-75, May, 1923.

A statement of the problem and our failure to meet it at present followed by an argument that individual differences must be recognized. It is implied that mental and subject-matter tests should be used for this purpose.

101. Miller, W. S. "General intelligence tests," School Review, 28:94, February, 1920.

A brief statement that the University of Minnesota High School has been classifying entering students according to mental test results with a list of favorable results. States that the highest correlation between mental test results and achievement is in mathematics and science.

102. Mitchell, David. "Psychological examination of pre-school age children," School and Society, 15:561-68, May 20, 1922.

Discusses the results of testing 1000 New York children, advocates special classes, differentiated curricula, etc.

103. Montgomery, E. W. "Group tests for intelligence in the Bedford (Indiana) schools," Sixth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 5, No. 1. Bloomington: Indiana University, 1919, p. 54-55.

The Binet-Simon Test was used to select pupils for special classes. In addition group tests were used in high school as the basis of organizing fast sections.

104. Murdoch, Katharine. "Rate of improvement of the feeble-minded as shown by standardized educational tests," Journal of Applied Psychology, 2:243-49, September, 1918.

The Stanford Revision and a dozen or so subject matter tests were used on a small number of feeble-minded children. Conclusion reached was that the mental age of a feeble-minded is not equivalent to the same mental age of a normal child.

105. Myers, A. F. "Reclassification of children on basis of tests in Port Clinton schools," Journal of Educational Method, 1:24-25, September, 1921.

In Port Clinton, Ohio, a change was made from the annual to the semester system by promoting the upper 40 percent of each grade one semester and demoting the lower 10 percent one semester. The basis of classification was a combined intelligence and achievement test score.

106. Myers, C. E., Myers, G. C. and Layton, S. H. "Group mental testing in Altoona, Pennsylvania," School and Society, 13:624-28, March 28, 1921.

The Myers' Mental Measure was used. It was recommended that special classes for sub- and super-normal children be formed, that several parallel courses be organized, that pupils be classified according to their I.Q.'s and school marks, etc.

107. Myers, G. C. "Economy in intelligence classification," Educational Administrtion and Supervision, 6:309-12, September, 1920.

A theoretical discussion of the grouping of pupils, maximum and minimum curricula, etc. Also an account of the testing and grouping of 1500 illiterate soldiers into four groups in each grade.

108. Myers, G. C. "Intelligence classification and mental hygiene," Pedagogical Seminary, 28:156-60, June, 1921.

An argument for complete mental surveys of school systems.

109. Nichols, M. L. "The teaching of science to classes divided according to ability." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania, Vol. 23, No. 1, p. 324-26.

In the South Philadelphia High School for Girls science classes were divided according to results of intelligence tests. Different methods of teaching were used in the different sections.

110. Omans, A. C. "Provision for ability grouping in junior and senior high school," American School Board Journal, 65:55-58, 138, October, 1922.

Gives questionnaire replies from fifty schools in all parts of the country. Shows that in general grouping on an ability basis is favored over individual promotions. Slightly over one-half classified on intelligence tests and teachers' estimates, about two-fifths on intelligence tests and school marks and a few on school marks alone. This is a good discussion and summary of practises.

111. PATTERSON, D. G. "A mental survey of the school population of a Kansas town," School and Society, 7:84-89, January 19, 1918.

This is well described by the title. A comparison is also made with Pintner's results in Ohio.

112. PINTNER, R. "A mental survey of the school population of a village," School and Society, 5:597-600, May 19, 1917.

An account of the survey of an Ohio village with a group of tests.

113. PINTNER, R. The Mental Survey. New York: D. Appleton and Company, 1918, 116 p.

This contains an account of using a number of intelligence and achievement tests, the results obtained from surveying school systems and the relation between educational accomplishments and mental ability, etc.

114. PINTNER, R. and CUNNINGHAM, BESS V. "The problem of group intelligence tests for very young children," Journal of Educational Psychology, 13:465-72, November, 1922.

A brief account of the organization of three sections in grade I of a school at Toledo, Ohio. The Pintner-Cunningham Primary Tests were used.

115. PINTNER, R. and MARSHALL, HELEN. "A combined mental-educational survey," Journal of Educational Psychology, 12:32-43, January, 1921.

A discussion of the necessity of combining the results of mental and educational tests. Also states that the schools have been best adapted to dull pupils.

116. PINTNER, R. and MARSHALL, HELEN. "Results of the combined mental-educational survey tests," Journal of Educational Psychology, 12:82-91, February, 1921.

A comparison of school work with capacity by means of mental and educational indices.

117. PINTNER, R. and NOBLE, H. "The classification of school children according to mental age," Journal of Educational Research, 2:713-28, November, 1920.

About 5000 pupils in a Columbus, Ohio, school were tested with the Stanford Revision. Fifteen percent were either given extra promotion or demoted. Special classes and an ungraded class were formed, acceleration and retardation were both reduced, discipline, work and interest were bettered.

118. PRESSEY, L. C. "The relation of intelligence to achievement in the second grade," Seventh Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 6, No. 1. Bloomington: Indiana University, 1920, p. 68-77.

Describe the use of the Indiana Primer Scale and Scale of Attainment No. I. in an Indiana city. It was found that no inferior pupils were high in achievement but that many superior pupils were low.

119. Pressey, S. L. "A systematic plan for selecting sub-normal and super-normal children in the public schools," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 92-99.

Recommends that a group test be used first and then the Stanford Revision.

120. Pressey, S. L. "The 'efficiency' of a group scale of intelligence in prognosticating success and failure in the junior high school," Journal of Applied Psychology, 3:381-85, December, 1919.

Discusses the use of the Indiana Group Point Scale in Bloomington. States that results are rather highly correlated with failure and success in school work.

121. Pressey, S. L. "School surveys by means of group tests of intelligence," Sixth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 5, No. 1. Bloomington: Indiana University, 1919, p. 46-53.

Discussion of the use of the Indiana Cross-Out Scale which is useful for general survey purposes.

122. Pressey, S. L. "Suggestions with regard to the use of mental tests and in particular with regard to their use in combination with tests of achievement," Seventh Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 6, No. 1. Bloomington: Indiana University, 1920, p. 78-80.

A discussion of general mental surveys, the use of achievement tests, their prognostic value, etc.

123. Pressey, S. L. "An attempt to measure the comparative importance of general intelligence and certain character traits in contributing to success in school," Elementary School Journal, 21:220-29, November, 1920.

A study of junior high-school pupils. The correlation of health, attitude, preparation and ability with school marks and with age was studied. Ability and attitude were found to be the most important.

124. Pressey, S. L. and L. W. "Measuring the 'usefulness' of tests in solving school problems," School and Society, 12:531-34, November 27, 1920.

The Indiana Cross-Out Scale was given to pupils in the upper grades and fast and slow sections arranged according to its results. Likewise the Primer Scale was given in the first grade. A better agreement was found between placement by test results and semester marks than between placement by teachers' judgments and semester marks.

125. PRICE, E. D. "The Enid plan of classification of pupils according to mental ability." Enid, Oklahoma: Board of Education, 1921. 12 p.

Elementary pupils were grouped into three groups according to teachers' opinions. The superior and inferior groups were also tested by the Terman Test, the results of which usually verified the teachers' judgment.

126. PROCTOR, W. M. "The use of intelligence tests in the educational guidance of high-school pupils," School and Society, 8:473-78, 502-09, October 19, 26, 1918.

In Palo Alto High School it was found that the Stanford Revision apparently affords as good a measure of success in school as do teachers' estimates or previous school marks. Pupils with I. Q.'s of 95 or below are not likely to be successful in high-school work.

127. PROCTOR, W. M. "Psychological tests and the probable school success of high-school pupils," Journal of Educational Research, 1:258-70, April, 1920.

A discussion of the use of the Stanford Revision and the Army Alpha Tests in high school and the connection between intelligence and high-school marks, elimination, retardation, etc. Army Alpha yields a rather low correlation with school marks.

128. PROCTOR, W. M. "The use of psychological tests in the educational guidance of high-school pupils," Journal of Educational Research, 1:369-81, May, 1920.

Shows that pupils guided in the choice of subjects according to Stanford Revision results show less elimination and failure than do those who are not so guided.

129. PROCTOR, W. M. "The use of psychological tests in the vocational guidance of high-school pupils," Journal of Educational Research, 2:533-46, September, 1920.

A discussion of the occupational levels of intelligence as found by testing in the army, the vocational choices of high-school pupils and the connection between the two. States that pupils should be guided out of vocations that require a greater degree of intelligence than they possess.

130. PROCTOR, W. M. and WARD, HELEN. "Relation of general intelligence to the persistence of educational and vocational plans of high-school pupils," Journal of Educational Research, 7:277-88, April, 1923.

Shows that there is a definite relation between results of intelligence tests and vocational plans and persistence and a still higher relation between the former and educational plans and persistence. Implies that groups of pupils formed according to intelligence test results should receive different treatment in our schools.

131. RACE, HENRIETTA V. "A study of a class of children of superior intelligence," Journal of Educational Psychology, 9:91-98, February, 1918.

In a Louisville school, pupils with I. Q.'s above 120 were selected. They did one year's work in one-half year without over twenty minutes of home study daily. The pupils appeared to be greatly benefited by this work.

132. Rapp, Anna A. "Grouping of children by abilities and consequent changes in school procedure. 2. Procedure in Reading." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 263-67.

The Terman and Illinois Tests were used in the seventh and eighth grades and pupils classified according to scores.

133. Reeve, W. D. "Homogeneous grouping of high-school students by means of psychological tests," Fourth Yearbook of the National Association of Secondary School Principals. Menasha, Wisconsin: George Banta Publishing Company, 1921, p. 81-94.

States that it is a waste of time to attempt to teach the best and worst 10 percent along with average pupils. Discusses the correlation of tests and school marks and says that classification by test results makes it easier to provide suitable instruction, reduces failures, gives the brighter pupils better training and in general conserves human resources.

134. Robbins, Chester. "The initial grouping of pupils." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 244-48.

A discussion of various articles in books dealing with this subject.

135. Rogers, Agnes L. "Mental tests as a means of selecting and classifying college students," Journal of Educational Psychology, 11:181-92, April, 1920.

Discusses results of testing about 300 Goucher college students with the Thorndike and Rogers Tests. Correlations with school marks were low for the single tests but high when the test results were combined. Shows that a division of students into three groups by test results would misplace about one-half, whereas by a chance placement two-thirds would be misplaced.

136. Rowland, S. V. "Individual difference among elementary grade pupils as evidenced by group intelligence tests." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 21, No. 37, Philadelphia, June 18, 1921, p. 101-04.

An account of classification based on results of the Stanford Revision and the Dearborn Tests. Special classes for inferior children were formed, also two or three groups of those remaining, and some individual instruction given.

137. Ruch, G. M. "An experiment with forced promotion," Educational Administration and Supervision, 6:71-73, February, 1920.

On the basis of several achievement tests a few pupils in the University of Oregon High School were skipped ahead a semester. Most of them made good.

138. Ruch, G. M. "Study of the mental, pedagogical and physical development of the junior division of the University High School, Eugene, Oregon." University of Oregon Publication, Vol. 1, No. 7, 1920.

A rather good discussion along the lines indicated.

139. Ruml, Beardsley. "Reliability of mental tests in the division of an academic group," Psychological Monographs, Vol. 24, No. 4, October, 1917.

A rather lengthy criticism of the use of marks as measures of ability and of the value of a number of tests for dividing pupils into instructional groups.

140. Saam, Theodore. "Intelligence testing as an aid to supervision," Elementary School Journal, 20:26-32, September, 1919. Addresses and Proceedings of the National Education Association, 57:625-29, 1919.

Twenty-three hundred and sixty lower-grade pupils at Council Bluffs, Iowa, were tested with the Stanford Revision. Kindergarten pupils were promoted to first grade on the basis of results. This procedure was satisfactory as a high correlation was found between I.Q.'s and school marks.

141. Schutte, T. H. "A mental survey in the training department of the Moorehead State Teachers College, Moorehead, Minnesota," American School Board Journal, 65:45-47, 53-54, 42-43; October, November, December, 1922.

The Otis, Chicago and Army Alpha Tests were given to elementary pupils and rapid progress groups formed in each grade. Most of the pupils in these groups gained one semester in the year.

142. Seashore, C. E. "Sectioning classes on the basis of ability," School and Society, 15:353-58, April 1, 1922.

Presents the following plan which was found successful in college. Two or three competitive exercises are set up at the first of the year and the marks thereon supplemented by mental tests and high-school marks as a basis of sectioning. Discusses advantages and objections.

143. Seashore, C. E. et al. "Mentality tests; a symposium," Journal of Applied Psychology, 7:229-40, 278-86, 348-60; April, May, June, 1916.

A general theoretical discussion of mental tests, their significance and use.

144. Shideler, J. W. "A correlation of teachers' grades and the scores of intelligence tests," School Review, 29:733-34, December, 1921.

An account of giving the Terman Group Test to 170 high-school pupils at Fort Scott, Kansas. Discusses its correlation with marks in the various subjects and finally concludes that intelligence test results should be used to supplement the teacher's judgment.

145. SPAIN, C. L. "Grouping of children by abilities and consequent changes in school procedure. 1. Procedure in Detroit." Ninth Annual Schoolmen's Week Proceedings, Vol. 23, No. 1, p. 257-63.

An account of the Detroit plan.

146. SPECHT, LOUISE F. "A Terman class in Public School No. 64, Manhattan," School and Society, 9:393-98, March 29, 1919.

Pupils in grades IVB to VIB were selected on the basis of low chronological age, high school marks, I.Q.'s of 120 or better, etc., and allowed to take an enriched curriculum and make individual progress. From one to four grades were covered in six months.

147. STEBBINS, R. and PECHSTEIN, L. A. "Quotients, I, E and A," Journal of Educational Psychology, 13:385-98, October, 1922.

A discussion of how to use test results and what should be the basis of classifying pupils.

148. STETSON, P. C. "Homogeneous grouping in the first year of a five-year high school," School Review, 29:351-65, May, 1921.

In the Muskegon, Michigan High School the Chicago Test was found unsatisfactory as a basis of classifying pupils but when supplemented by teachers' opinions gave better results. Both teachers and pupils favored having three groups on the basis of ability.

149. TAYLOR, J. S. "Grading and promotion," School and Society, 17:405-09, April 14, 1923.

Tells of classification in one New York City district which was based upon I.Q.'s, A.Q.'s, teachers' estimates, etc.

150. TERMAN, L. M. "Intelligence tests as a basis for grading." The Measurement of Intelligence. Boston: Houghton Mifflin Company, 1916, p. 16-17.

A brief statement that promotions should be made chiefly on the basis of intellectual ability.

151. TERMAN, L. M. "The use of intelligence tests in the army," Psychological Bulletin, No. 15, p. 177-87, June, 1918.

A good general account of the army work.

152. TERMAN, L. M. The Intelligence of School Children. Boston: Houghton Mifflin Company, 1919, 317 p.

A discussion of many phases of intelligence testing including general principles, the amount of individual differences found, capacity according to mental age, prediction according to the I.Q., etc.

153. TERMAN, L. M. "The use of intelligence tests in the grading of school children," Journal of Educational Research, 1:20-32, January, 1920.

Mostly a theoretical discussion of individual differences, how many pupils should be accelerated and how many retarded, etc. Summarizes what has been found as to difference of test results and teachers' opinions as showing that tests are usually more nearly right.

154. Terman, L. M. "The psychological determinist; or democracy and the I.Q.," Journal of Educational Research, 6:57-62, June 1922.

An answer to Bagley's speech against the use of the I.Q. for classifying pupils.

155. Terman, L. M. et al. Intelligence Tests and School Reorganization. Yonkers, New York: World Book Company, 1922, 111 p.

A theoretical discussion of the problems accompanied by statements of what is being done in Oakland, Los Angeles and Miami, Arizona.

156. Thorndike, E. L. et al. "Intelligence and its measurement: A symposium," Journal of Educational Psychology, 12:123-47, 195-216, 271-75; March, April, May, 1921.

A rather theoretical discussion of the nature and measurement of intelligence and the next steps in research dealing with it.

157. Tildsey, J. L. "Some possibilities arising from the use of intelligence tests," Bulletin of High Points in the Work of the High Schools of New York City, 3:4-10, June, 1921.

Tells of various uses of intelligence tests in New York.

158. Trabue, M. R. "Some pitfalls in the administrative use of intelligence tests," Journal of Educational Research, 6:1-11, June, 1922.

A critical discussion pointing out certain dangers to be avoided.

159. Trabue, M. R. and Stockbridge, F. P. "Psychological tests in education," Chapter VII, Measure Your Mind. New York: Doubleday Page and Company, 1920, p. 63-75.

Among other things contains a brief account of an experiment in Public School No. 64, New York City, wherein mental tests were one of the bases used to select pupils for rapid advancement. Also discusses the predictive value of mental tests in regard to college marks and shows that it is rather high.

160. Tupper, C. R. "The grading of pupils in Miami, Arizona," Educational Research Bulletin, 2:163-65, May 30, 1923.

A description of the plan in use in Miami in which class groupings are made more or less on the basis of mental age. The teachers appear to favor the plan. It has reduced failure by more than two-thirds and increased acceleration considerably.

161. Wallin, J. E. W. "A comparison of three methods for making the initial selection of presumptive mental defectives," School and Society, 13:31-45, January, 1921.

States that the Pressey Primer Scale is inferior to the judgment of trained individuals in selecting defective individuals.

162. WASHBURNE, C. W. "The individual system in Winnetka," Elementary School Journal, 21:52-68, September, 1920.

Describes the system in Winnetka in which progress and promotion are entirely individual, in which most of the work is done by using practise material and then taking complete diagnostic tests. States that every teacher favors the plan, that at least five percent were saved from repeating and $5000 saved in one year.

163. WASHBURNE, C. W. "Educational measurements as a key to individual instruction and promotions," Journal of Educational Research, 5:195-206, March, 1922.

An account of the individual system used in Winnetka.

164. WHIPPLE, G. M. "The problem of selecting and training gifted children in the public schools," Fifth Conference on Educational Measurements, Bulletin of the Extension Division, Vol. 4, No. 4. Bloomington: Indiana University, 1918, p. 6-25.

An argument for the selection of gifted children by mental tests and the providing of special opportunities for them. States that the 10 percent with I.Q.'s above 115 can probably do two years' work in one. Also discusses the Urbana and other experiments.

165. WHIPPLE, G. M. "Educational determinism; a discussion of Professor Bagley's address at Chicago," School and Society, 15:599-602, June 3, 1922.

A strong reply to Bagley's address in which the validity and use of intelligence tests are strongly advocated.

166. WHITCOMB, M. EDITH. "Intelligence tests in the primary grades," Journal of Educational Research, 5:58-61, January, 1922.

An account of the work done with the Stanford Revision in Council Bluffs, Iowa. Gives various data and shows that the plan of placing pupils by test scores works well.

167. WHITMIRE, ETHEL D. "Intelligence tests vs. teacher's estimates," Psychological Clinic, 13:197-98, May 15, 1920.

Discusses a couple of cases in which teachers rated pupils by their personality and attitude.

168. WHITNEY, F. P. "Provision for accelerant and retarded children in junior high school," School Review, 27:695-705, December, 1919.

In the Collinswood Junior High School of Cleveland pupils were placed by school marks and test results.

169. WILLETT, G. W. "A suggestion for meeting individual differences," School Review, 28:576-84, October, 1920.

States that cumulative records are better than either teachers' judgments or intelligence test results, but that a combination of all three is best.

170. WILLIAMS, A. J. "Age-grade distributions and intelligence quotients," Journal of Educational Psychology, 11:39-44, January, 1920.

An account of the Otis Test in Brewster and a discussion of its results as compared with the age-grade location of the pupils.

171. WILLIS, C. B. "The grading and promotion of pupils," Journal of Educational Method, 1:90-95, November, 1921.

An argument for the classification of pupils according to their ability.

172. WINTER, O. "Chicago intelligence test in Harrison Technical High School," School Review, 28:772-75, December, 1920.

It was found that the correlation between intelligence test scores and certain achievement tests and also school marks was rather low.

173. WOODY, CLIFFORD. "Tests and measures in the schoolroom and their value to the teachers," School and Society, 6:61-66, July 21, 1917.

Mentions others but especially tells of an experiment in Chatham, New Jersey. On the basis of scores on five achievement tests the best pupils were given extra promotion. All made good and led in the classes in which they were placed.

174. WOODY, CLIFFORD. "Measurement of the effectiveness of differentiation of high-school pupils on the basis of the Army intelligence tests," Journal of Educational Research, 7:397-409, May, 1923.

A study of the freshman class in a small Michigan high school. Shows that if the class had been divided into three groups according to scores upon Army Alpha the division would have agreed fairly well with achievement as measured by tests in English, Algebra and Latin.

175. WURTH, CARL. "Intelligence tests and promotions," American School Board Journal, 65:50-51, December, 1922.

An argument against segregating bright pupils for three reasons. They should spend their time getting work better rather than going more rapidly, their absence reduces the enthusiasm of the class and they acquire vanity and false pride from being classified as superior.

176. YATES, DOROTHY H. "Impressions from two years spent in testing gifted children." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers' College. Sacramento, California: State Printing Office, 1923, p. 29-33.

A discussion of gifted children with some argument to the point that training for leadership does not need to be begun early and that bright pupils should not be forced too much.

177. Yoakum, C. S. and Yerkes, R. M. Army Mental Tests. New York: Henry Holt and Company, 1920, 303 p.

A general account of how the Army Tests were made and used with a discussion of results and applications.

178. Zirkle, H. W. "Character and results of special rooms as conducted in the Whittier School of Denver," Elementary School Journal, 21:189-97, November, 1920.

Pupils are placed in regular, accelerated and retarded rooms according to test results, age, school marks, health, etc.

179. Zirkle, H. W. "Taking care of the gifted child," The Technique of Supervision by the Elementary School Principal. The First Yearbook of the Department of Elementary School Principals. Washington: National Education Association, May, 1922, p. 81-86.

In a Denver school, superior pupils are selected by individual and group tests, school marks, effort and physical condition and are given the opportunity to gain time. Practically all of the pupils selected are able to gain time and still continue to make high school marks.

180. Zornow, T. A. and Peckstein, L. A. "An experiment in the classification of first-grade children through the use of mental tests," Elementary School Journal, 23: 136-46, October, 1922.

In Rochester, New York, over 400 entering pupils were tested with the Stanford Revision and classified according to the test results. Later the Detroit First-Grade Group Test was used. It was found that but few pupils had to be shifted after being thus placed.

181. "Classification of pupils according to mentality," The School Magazine, 3:196-97, April, 1921.

In a Buffalo school A, B and C groups were formed in grades VI-VIII according to the average of Otis scores, teachers' marks and an average mark. The plan appears to have been successful.

182. "Classification of pupils in Miami, Arizona," Elementary School Journal, 22:727-29, June, 1922.

When pupils were classified on the basis of mental tests, the number of failures was reduced and much money saved.

183. "Double promotion," Elementary School Journal, 23:324-25, January, 1923. An editorial.

In Macomb, Mississippi, all pupils who are recommended by the teacher and have a doctor's certificate of good health, who are not below a certain age, whose previous record in school work and attendance is satisfactory and who make certain scores upon achievement tests are given double promotion if it does not involve skipping certain subjects.

184. Educational Research Bulletin, Ohio State University, Vol. 2, No. 8, p. 123-24, April 18, 1923.

An account of procedure in Painesville, Ohio. In grades VIIB to VIIIA pupils are grouped according to the amount of work they are capable of doing but in the lower grades they are grouped according to ability as shown by the Illinois Examination and school work. Also there is a special room with individual progress and promotion upon effort for upper-grade pupils whose school work is unsatisfactory and whose ability appears to be low.

185. "Grouping pupils on the basis of ability." The First of a Series of Surveys of the Dept. of Instruction, Cleveland Public Schools, Board of Education, Cleveland, Ohio. p. 18-21.

In the first and seventh grades pupils were classified chiefly according to results from the Pressey Primer and Illinois Scales, respectively.

186. "Intelligence tests as a basis for homogeneous grouping at Xenia, Ohio," American School Board Journal, 64:75-76, March, 1922.

Three groups were made up of seventh and eighth grade pupils on the basis of the Haggerty and Terman Tests. They were regrouped later as was necessary but in most cases stayed in the same group.

187. Journal of Educational Research, 7:265, March, 1923.

In Painesville, Ohio, seventh and eighth grade pupils are grouped according to the Illinois Examination results. The sub-normal of these grades are promoted on the basis of their effort.

188. Journal of Educational Research, 7:454, May, 1923.

A brief account of the reclassification of pupils in grades V - VIII on the basis of scores upon the National Group Intelligence Test, teachers' marks and the time devoted to home work. Such changes in the original classification as seemed necessary were made. The plan seems to have been successful.

189. Journal of Educational Research, 7:456-57, May, 1923.

In Miami, Arizona, results from the Otis and National Group Intelligence Tests are used to determine the I.Q.'s of the pupils for the purpose of grouping. Each teacher handles two sections of different ability, the average and the dull. Extra promotions are given to take care of the brighter pupils. The failure rate has been decreased very markedly while results of standardized tests show that achievement has not been lowered.

190. "Provision for exceptional pupils in Oakland," School and Society, 14:30-31, July 16, 1921.

Gives data as to the number of special classes and special promotions for exceptional pupils.

191. "Pupils classified by mental tests," School Life, 12:13, February 1, 1921.

An account of reclassification of pupils in the W. T. Harris school of New York City by group tests.

192. "Reclassified school making good progress," School Life, May 1, 1921. p. 12.

A brief account of the working of the plan in the W. T. Harris school of New York City in which pupils were grouped in sections by group test results. The plan appears to be working satisfactorily.

193. "Report of the school committee for the year 1917-1918." Providence, Rhode Island: School Committee, p. 49-70.

Discusses many points in connection with the use of intelligence tests, especially their high correlation with quality of school work and their reliability for classifying pupils.

194. "Uniform classification by mental age," School Life, 8:88, December, 1921.

A statement that in Denver it is planned to reclassify all pupils into homogeneous groups according to their mental ages.

PART II.

195. ALDERMAN, L. R. "An effort to make the school fit the needs of the exceptional child," Addresses and Proceedings of the National Education Association, 52:830-35, 1914.

An account of provisions for exceptional children in Portland, Oregon, followed by general discussion. Provisions include a special school for defectives and special rooms for other exceptional children.

196. ALEY, R. J. "Care of exceptional children in the grades," Addresses and Proceedings of the National Education Association, 48:881-86, 1910.

A general discussion, its chief point being that superior pupils have been comparatively neglected.

197. APPELL, I. and WOLFSON, A. N. "Plan for organizing the entrants into the high school of commerce according to their attainments in scholarship, their capacities and their aims," School Review, 27:256-61, April, 1919.

In the New York High School of Commerce pupils were formerly grouped according to the foreign language they had studied but the result was not satisfactory. Later they were divided for English work according to their marks in English during the first three weeks, similarly for mathematics. Results were fairly satisfactory.

198. BAGLEY, W. C. "The 'Batavia system' of class-individual instruction," Classroom Management. New York: Macmillan Company, 1916, p. 214-24.

A description of the "Batavia system."

199. Barnard, F. J. "Classification and promotion of pupils." Report of United States Commissioner of Education, Vol. 1, p. 303-56, 1898-99.

An account of a number of flexible plans of promotion including those in operation in St. Louis, Elizabeth, Seattle, Denver, and Cambridge.

200. Barnard, F. J. "Let pupils be so classified as to allow unrestricted progress or unlimited time, according to ability," Addresses and Proceedings of the National Education Association, 38:163-70, 1899.

An account of the Seattle plan after it had been in operation eight years. Data are given for over 7000 pupils enrolled in four parallel tracks. The plan appears to have been successful.

201. Bates, W. C. "Annual Report of the School Committee and the Superintendent of Schools." Cambridge, Massachusetts, 1908, p. 45-51.

A description of the Cambridge double-track plan. Data are given for seventeen years' operation showing that on the whole the more rapidly pupils progressed through the grades the better marks they made in high school.

202. Becht, J. G. "Bright pupils and dull pupils," Journal of Education, 79:395-6, April 9, 1914.

A discussion and diagnosis of the causes of retardation and of the fact that bright pupils deserve special help.

203. Belisle, H. J. "An experiment in retardation," Journal of Education, 75:381, 87-9, April 4, 1912.

An account of the organization of fast, medium and slow classes in Lawrence, Massachusetts. The operation of the plan appears to have been satisfactory.

204. Bergen, J. Y. "Grading inside of class lines," Educational Review, 16:81-85, June, 1898.

Description of the division of freshmen and other students of the English High School of Boston according to their average marks in various subjects.

205. Berry, C. S. "Special classes in Michigan for mentally exceptional children." Report of Superintendent of Public Instruction, Michigan, Vol. 78, 1914-15, p. 57-68.

Gives the results obtained from a questionnaire answered by over 200 towns. Five different classes of exceptional children are considered. Fifty-eight percent of school administrators favor classes for bright children.

206. Blewett, Ben. "The system of grading pupils in St. Louis," Educational Review, 8:387-8, November, 1894.

An explanation of the St. Louis quarterly system.

207. BOYKIN, J. C. "Class intervals in city public schools." Report of the United States Commissioner of Education, Vol. 2, p. 981-1009, 1891.

A very good sketch of the history of graded schools from the sixteenth century to the time of writing. The St. Louis and other flexible systems are described.

208. BREHM, J. J. "Variation in the treatment of different groups." Ninth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 23, No. 1, p. 248-52.

In the junior high school of Harrisburg pupils are grouped according to their previous records. The school is "thoroughly committed to this plan."

209. BROWN, S. W. "Some experiments in elementary school organization," Addresses and Proceedings of the National Education Association, 51:458-63, 1913.

The chief point made is an argument for individual progress in each subject at the rate best suited to the pupil.

210. BUCHANAN, ELIZABETH. "Classification and promotion," Addresses and Proceedings of the National Education Association, 39:128-37, 1900.

An argument for short class intervals, of not more than twelve weeks, or even six.

211. BURK, C. F. "Promotion of bright and slow children," Educational Review, 19:296-302, March, 1900.

A description of the Santa Barbara system with A, B, and C sections in each grade. These sections differed in the intensity and extensity of work covered.

212. BURK, FREDERICK. "Every child a minor vs. the lockstep. A suit in equity." Monograph C. San Francisco State Normal School, Fresno, 1915.

A description of individual instruction in the training school for practice teachers. Each pupil goes at his own pace and promotion is made at any time.

213. BURNELL, ELIZABETH F. "Instruction in mathematics for gifted pupils," Pedagogical Seminary, 24:569-83, December, 1917.

Gives the results of a questionnaire on provisions for bright children. Also a bibliography.

214. BUSWELL, G. T. "The school treatment of mentally exceptional children." Elementary School Journal, 23:683-93, May, 1923.

A brief discussion of a number of plans of providing for mentally exceptional children. The argument favors modification of the curriculum and varying the methods of work rather than varying the rate of progress.

215. CARFREY, J. H. "Grading and promotions," Journal of Education, 75:379-80, April 4, 1912.

States that shorter intervals than semesters are best. In Franklin, Massachusetts, a modification of the shorter interval and the group plans are used.

216. Chalmers, W. W. et al. "Promotions and gradings," Addresses and Proceedings of the National Education Association, 38:369-75, 1899.

A general discussion of semester vs. annual promotions, various flexible plans of organization, etc.

217. Chalmers, W. S. "Semi-annual promotions." Annual Report of the Board of Education of the City of Toledo, Ohio, for the school year ending August 31, 1903. Toledo, Ohio: Board of Education, 1903, p. 30-41.

An argument showing the importance of semi-annual promotions as a device for getting away from the annual "lockstep." Contains long quotations from W. H. Harris.

218. Clerk, F. E. "The Arlington plan of grouping pupils according to ability in the Arlington High School, Arlington, Massachusetts. School Review, 25:26-48, January, 1917.

A description of the Arlington, Massachusetts plan which relocates pupils every two months on the basis of school marks. Nine advantages are stated and eleven objections answered on the basis of six years of experience with the plan.

219. Cogswell, Francis. "The Cambridge experiment," Addresses and Proceedings of the National Education Association, 33:333-38, 1894.

A description of the Cambridge, Massachusetts, plan which provides for a four-and a six-year track through grades four to nine inclusive.

220. Coleman, E. M. "Grading for efficient organization in the interests of pupils," Addresses and Proceedings of the National Education Association, 40:286-7, 1901.

Describes the plan in use at Ft. Dodge, Iowa, which was adapted from the St. Louis plan.

221. Collicott, J. G. "The bright pupil," Addresses and Proceedings of the National Education Association, 53:457-66, 1915.

Gives results from a questionnaire sent out to all cities of over 50,000. Discusses briefly about a dozen flexible plans for the selection of bright pupils. Followed by arguments for and against.

222. Corson, D. B. "Classification of pupils," Educational Administration and Supervision, 6:86-93, February, 1920.

An argument for rapid, regular and slow sections of pupils.

223. Cubberley, E. P. "Promotional plans." Public School Administration. Boston: Houghton Mifflin Company, 1916, p. 300-06.

Describes the Pueblo, Batavia, Cambridge, Mannheim, and a few other plans.

224. CUMMINS, R. A. "'Bright' and 'slow' pupils in elementary and high school," Journal of Educational Psychology, 10:377-88, October, 1919.

An account of a study in a private school in New York City. It was found that there was a tendency for pupils who were bright and making fast progress to continue to do so and for others to continue to make slow progress.

225. DEMPSEY, C. H. "Flexible grading and promotions," Journal of Education, 75:373-5, April 4, 1912.

Describes the reorganization of the system at Malden, Massachusetts, which provided opportunity for bright pupils to skip three semesters of work. Lists the principles upon which reorganization was based.

226. DE VOSS, J. C. "School success of gifted children." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 34-38.

An argument to the effect that gifted children are able to carry extra work and progress more rapidly. To some extent replies to Dr. Bagley's address on educational determinism.

227. DOUGHERTY, N. C. "Report of the round-table discussion on 'promotion in city schools'," Addresses and Proceedings of the National Education Association, 31:802-3, 1892.

States that in Chicago promotions are made at all times, only slightly over half being at the end of the year. Argues that a flexible system helps pupils.

228. DOWNES, F. E. "Seven years with unusually gifted pupils," Psychological Clinic, 6:13-17, March 15, 1912.

Describes the bright sections organized in the Harrisburg, Pennsylvania, schools, also the special schools for bright pupils of the upper grades. States that attendance, scholarship, discipline and high-school records were all improved by the operation of the plans described.

229. DOWNES, F. E. "Other special schools." Report of the Superintendent of Schools, Harrisburg Public Schools, 1912, p. 23-26.

Describes the schools for exceptionally gifted children which allow them to cover work of the eighth and ninth years in one year. States advantages of this plan and that the pupils involved do well in high school.

230. DUTTON, S. T. and SNEDDEN, DAVID. "Grading and promotion." The Administration of Public Education in the United States, New York: Macmillan Company, 1912, p. 341-55.

A discussion of homogeneity within a grade, reclassification and flexible grading, minimum and maximum requirements, individual teaching, ungraded rooms, the Cambridge and Batavia plans, etc.

231. DYER, F. B. "Superintendent's report." Seventy-ninth Annual Report of the Public Schools of Cincinnati, 1908, p. 66, 67-69.

A brief account of certain special classes and vacation schools in Cincinnati.

232. Dyer, F. B. "Rapid advancement classes." Annual Report of the Superintendent, Boston, Public Schools, 1913, p. 65-67.

An account of certain classes in which pupils may do three years of work in two. The plan worked well and was to be extended to other schools of the city.

233. Fisher, G. M. "Report to superintendent." Sixty-fourth Annual Report of Schools of Worcester, 1912, p. 74-75.

An account of the summer school provided for failing pupils to make up work and for bright pupils to gain time.

234. Freeman, F. N. "Provision in the elementary school for superior children," Elementary School Journal, 21:117-31, October, 1920.

Gives results of a questionnaire sent out to cities of over 25,000. A rather careful analysis of results is made showing what plans are in use and upon what they are based. Advantages and difficulties are discussed.

235. Garber, J. P. "A rational system of classification and promotion of pupils in elementary schools," Education, 27:288-302, January, 1907.

Argues that promotional intervals should fit pupils but that bright and dull pupils should not be separated nor should intervals be very short. Also says that failure is not the chief cause of elimination.

236. Gillingham, Anna. "The bright child and the school," Journal of Educational Psychology, 10:237-52, May-June, 1919.

A description of a number of bright children and an argument that such children be given more recreation and play rather than extra work leading to accelerated progress.

237. Gist, A. S. "The acceleration of pupils," School and Society, 5:116-18, January 27, 1917.

An account of an experiment in grouping pupils according to teachers' estimates in a Seattle school. The experiment was considered a success.

238. Glass, J. M. "The study-coach or opportunity class organization." Eighth Annual Schoolmen's Week Proceedings, University of Pennsylvania Bulletin, Vol. 21, No. 37, Philadelphia, June 18, 1921, p. 250-56.

Describes two types of study-coach organization to meet the individual needs of three classes of pupils. Gives a list of advantages of so doing.

239. Gordon, C. H. "Reorganization of the grammar school and a rational system of grading," Education, 21:16-26, September, 1900.

An argument for two parallel courses of the same length, one of which has a richer curriculum than the other, and for the basing of promotion chiefly upon effort.

240. Gosling, T. W. "The classification of pupils," Educational Review, 39:394-399, April, 1910.

A discussion of the problem as it exists in high school.

241. Greenwood, J. M. "Shorter time in elementary school work," Educational Review, 24:375-90, November, 1902.

Gives data concerning the Kansas City elementary schools which do in seven years what most schools do in eight. Shows that the work is satisfactorily done.

242. Gregory, Christopher. "Holding pupils in school," Atlantic Educational Journal, 11:436-37, April, 1916.

Describes the semi-annual plan with special classes for retarded and rapid classes for superior pupils in Long Branch, New Jersey. States that it results in the seventh and eighth grades being as large as the fifth and sixth.

243. Harley, H. L. "The physical status of the special class for bright children in the University of Pennsylvania, summer session of 1912," Psychological Clinic, 7:20-23, March 15, 1913.

Shows that school work done in hot weather did not injure the health of bright pupils. All except one increased in weight while in summer school.

244. Harris, W. T. "Advantages and disadvantages of the graded system." Fifteenth Annual Report of the St. Louis Public Schools, 1868-69, p. 104-07.

This discussion implies dissatisfaction with the rigid grading system.

245. Harris, W. T. "Superintendent's report." Eighteenth Annual Report of the St. Louis Public Schools, 1871-72, p. 24-27, 81-87.

Mentions the quarterly promotions in use in St. Louis, and discusses the advantages of classification if not too rigid. States that a five-weeks interval is ideal and that a system must be readjusted to meet the needs of its pupils.

246. Harris, W. T. "Superintendent's report." Nineteenth Annual Report of the St. Louis Public Schools, 1872-73, p. 24-29.

An explanation of the flexible system used in St. Louis which provided for regular quarterly promotions but also had some groups organized with only five weeks intervals.

247. Harris, W. T. "Superintendent's report." Twentieth Annual Report of the St. Louis Public Schools, 1873-74, p. 121-48.

An argument in favor of the St. Louis and other flexible plans of organization. Contains quotations from several others in favor of such plans.

248. Harris, W. T. "Superintendent's report." Twenty-first Annual Report of the St. Louis Public Schools, 1874-75, p. 28-29, 86-93.

Contains quotations from other exponents of flexibility.

249. HARRIS, W. T. "Classification in graded schools." Report of the United States Commissioner of Education, Vol 1, p. 601-36, 1891-92.

This is merely a collection of what is contained in the various reports of the St. Louis Public Schools on the same subject.

250. HARRIS, W. T. "Class intervals in graded schools," Addresses and Proceedings of the National Education Association, 39:332-40, 1900.

A statement of the need of flexible organization and of separating weak and strong pupils. It is also argued that promotion should be individual and not by classes.

251. HARRIS, W. T. et al. "The early withdrawal of pupils from school: its causes and its remedies," Addresses and Proceedings of the National Education Association, 12:260-73, 1872.

Several causes of elimination are stated of which defective grading is the most potent. Discussion both for and against short intervals of organization follows.

252. HARTWELL, C. S. "Economy in education," Educational Review, 30:159-77, September, 1905.

States that the grammar school is not satisfactory and that promotion by subjects is one of the changes needed.

253. HARTWELL, C. S. "The grading and promotion of pupils," Educational Review, 40:375-86, November, 1910. Addresses and Proceedings of the National Education Association, 48:294-306, 1910.

An argument for flexible organization and promotion by subjects. The results of a questionnaire are given showing more favorable sentiment on these points than existed several years previously. It is also stated that bright and dull pupils should be given individual attention.

254. HATCH, W. E. et al. "Provisions for exceptional children in the public schools," Addresses and Proceedings of the National Education Association, 45:360-63, 1907.

Rather brief discussion of several of the plans.

255. HEILMAN, J. D. "The need for special classes in the public schools," Psychological Clinic, 1:104-14, June 15, 1907.

This deals entirely with classes for inferior pupils.

256. HENRY, T. S. "Classroom problems in the education of gifted children." Nineteenth Yearbook of the National Society for the Study of Education, Part II, 1920. 125 p.

A complete account of an experiment in selecting gifted children at Urbana, Ill., and a brief description of a number of flexible promotion schemes. There is also a bibliography on the psychology and pedagogy of gifted children.

257. Hoblit, M. L. "The high-school unit: quantity, quality, and credit," School Review, 23:303-06, May, 1915.

A suggestion that high-school graduation be determined by a combined system of units and points, the latter to depend upon the marks made. The scheme is so arranged that a superior student can complete the work in three years, an average student in four years and a poor student in more than four.

258. Hoblit, M. L. "Is credit for quality sound?—A rejoinder," School Review, 23:712-14, December, 1915.

States that his suggested plan did not involve the graduation of a pupil who had covered less than fifteen units of work. Also argues in its defense.

259. Holmes, W. H. "Plans of classification in the public schools," Pedagogical Seminary, 18:475-522, December, 1911.

Describes a dozen or more of the different plans used in this country, also several used in England. The discussion is followed by a bibliography.

260. Holmes, W. H. "The St. Louis Plan," Journal of Education, 75:380, April 4, 1912.

A brief description of the plan mentioned.

261. Holmes, W. H. "School organization and the individual child." Worcester: Davis Press, 1912, Partly in Journal of Education, 75:376-9, April 4, 1912.

A long discussion of a number of different plans of classification and promotion. Also a theoretical discussion of the subject. Some space is given to mental tests.

262. Hutton, T. B. "Classification and gradation," School Review, 16:543-50, October, 1908.

Describes the two-group system in use at LeMars, Iowa. States that it prevents retardation and saves money.

263. Jackman, W. S. "The school grade a fiction," Educational Review, 15:456-73, May, 1898.

A strong plea and argument against having eight rigidly defined elementary grades.

264. Johnson, F. W. "Credit for courses in the University High School," School Review, 23:715-18, December, 1915

Gives the plan of credit for quality followed in the University of Chicago High School. The credit ranges from .85 for a grade of 60 percent to 1.25 for a grade of 95 percent. States that the University of Chicago admits some students on this basis.

265. Jones, E. E. "A concrete example of the value of individual teaching," Psychological Clinic, 2:195-203, December 15, 1908.

Gives the history of one successful case and states a few general principles.

266. JONES, E. E. "Suggestions from cases of unusually rapid or irregular progress in public schools," Addresses and Proceedings of the National Education Association, 50:640-45, 1912.

A study of several hundred pupils who had skipped grades which shows that most of them should have skipped and that a few should not.

267. JONES, H. S. et al. "II—Report of the committee on city school systems. Pupils-classification, examination and promotion." Addresses and Proceedings of the National Education Association, 26:276-84, 1886.

In this discussion are mentioned various types of classification and the need for considering individual pupils.

268. JONES, OLIVE M. Teaching Children to Study. The Group System Applied. New York: Macmillan Company, 1909.

Gives a brief account of a number of plans and a very full one of the Mannheim special class system.

269. JONES, W. F. "An experimental-critical study of the problem of grading and promotion," Psychological Clinic, 5:63-96, 99-120, May 15, and June 15, 1911.

A discussion of a number of double-track systems and of the reasons for failure. States that the pupils' and teachers' reasons for the latter are as often false as true.

270. KENDALL, C. N. "What modifications in organization are necessary to secure suitable recognition for pupils of varying ability, particularly for the ablest?" Addresses and Proceedings of the National Education Association, 46:147-55, 1908.

Discusses the Cambridge and other plans. Is in favor of making some modifications.

271. KENNEDY, JOHN. "The Batavia plan after fourteen years of trial," Elementary School Teacher, 12:449-59, June, 1912.

After fourteen years experience with the Batavia plan the author is still an enthusiastic believer in its advantages.

272. KENNEDY, JOHN et al. The Batavia System of Individual Instruction. Syracuse: C. W. Bardeen, 1914. 299 p.

A complete description of the Batavia system, with two teachers in each room, one the regular teacher and the other to aid individuals. States that elimination and expense were reduced, teachers freed from worry and the general school spirit bettered.

273. KENNEDY, J. W. "The all-year school," Addresses and Proceedings of the National Education Association, 55:795-801, 1917.

Tells of summer term and "promotion classes" in Newark. Both time and money were saved.

274. KILPATRICK, V. E. "Emancipating the individual pupil," Education, 30:375-85, February, 1910.

A general discussion of individuality followed by an enumeration of ten plans of school organization.

275. KIRK, J. R. "Should the school furnish better training for the non-average child?" Addresses and Proceedings of the National Education Association, 45:221-27, 1907.

A plea for flexible organization.

276. LEIGHTON, E. V. "At last—a chance for the gifted child," Popular Educator, 30:115, October, 1912.

Brief mention of several cities making provision for or giving recognition to gifted children.

277. LEWIS, H. P. "Semi-annual promotion." Report of the Public Schools of the City of Worcester. Worcester, Massachusetts: School Committee, 1904, p. 34-41.

An argument in favor of semi-annual promotions. Contains quotations from several sources.

278. LITTWIN, M. F. "English 'star' classes at the Boys' High School." Bulletin of High Points in the Work of the High Schools of New York City, 5:4-7, March, 1923.

On the basis of school marks special sections were formed in the sixth, seventh and eighth semesters of high school. These sections were given differentiated and enriched work in English. A number of advantages were derived from the plan.

279. LYNCH, ELLA F. "The bright child," Psychological Clinic, 4:141-44, October 15, 1910.

A plea that the bright child be given individual instruction and the opportunity for individual progress.

280. MCDONALD, R. A. F. "Provision for the exceptionally gifted." Adjustment of School Organization to Various Population Groups. Teachers College Contributions to Education, No. 75. New York: Teachers College Bureau of Publications, p. 90-101.

A theoretical discussion followed by a brief description of about a dozen plans.

281. MAENNEL, B. "The auxiliary schools of Germany." (Translated by Dresslar, F. B.). United States Bureau of Education Bulletin, No. 3, 1907, p. 43-47, 121-23.

An account of the Mannheim three-fold plan.

282. MANGOLD, G. B. "The mental classification of children. The education of exceptional children." Problems of Child Welfare. New York: Macmillan Company, 1914, p. 205-27.

This is almost entirely a discussion of backward children.

283. MEYER, M. F. "Is credit for quality sound?—A criticism," School Review, 23:708-11, December, 1915.

An argument in favor of credit for quality which is in use in a portion of the University of Missouri. Especially answers Professor Ruediger's adverse criticism.

284. MILLER, C. A. A. J. "The study of exceptional children," Addresses and Proceednigs of the National Education Association, 46:957-63, 1908.

A theoretical discussion listing ten kinds of exceptional children.

285. MITCHELL, DAVID. "Schools and classes for exceptional children." Educational Survey. Cleveland, Ohio: Cleveland Foundation, 1916. 122 p.

Discusses provisions for ten classes of exceptional children in Cleveland. Bright children are not included in those provided for.

286. MULREY, CORA L. "The rapid advancement class," Educational Administration and Supervision, 3:416-19, September, 1917

A brief account of the Cambridge plan by which three years of work may be done in two.

287. MYERS, G. C. "Broadening the course of study for the bright child," Educational Administration and Supervision, 3:33-37, January, 1917.

Suggests minimum and maximum courses and answers a number of objections to them.

288. NEVERMAN, P. F. "New Richmond plan of grade promotion," American School Board Journal, 54:38, January, 1917.

Describes the New Richmond, Wisconsin, plan which provided two tracks through the lower grades. Also a number of principles upon which the plan is based are given.

289. PARLIN, F. E. "Gradation and promotion." Annual Report of the School Committee and the Superintendent of Schools, Cambridge, Massachusetts, 1910, p. 19-21.

A brief explanation of the Cambridge plan.

290. PARKINSON, W. D. "Promotions, accelerated and retarded," Education, 19:152-57, November, 1898.

An argument that there is some advantage in the inertia of the graded system and that only a comparatively few pupils should be accelerated and retarded.

291. PARKINSON, W. D. "Individuality and social adjustment as means and ends in education," Education, 29:16-24, 104-12, September, October, 1908.

A discussion of individual and class instruction and progress. Argues for a fairly flexible organization adapted to the individual.

292. PATTERSON, M. ROSE. "A preparatory center in Baltimore," Atlantic Educational Journal, 12:234-38, January, 1917.

A description of the Baltimore plan in which the best upper grade pupils of the city are brought together at the end of the sixth grade and may either do three years of work in two or two in one and one-half.

293. PATTERSON, M. ROSE. "The Homewood demonstration school at Johns Hopkins University," School and Society, 16:577-84, November 18, 1922.

An account of an experiment in which a number of pupils did five months work in two. Especially the under-age pupils succeeded in doing the extra work.

294. PAYNE, W. H. "Elastic grading." Report of the United States Commissioner of Education, Vol. 2, p. 1376-77, 1899-1900.

Argues that the best basis of classification is the average of the year's work and that on such a basis few pupils should need reclassification.

295. PHILLIPS, D. E. "The child vs. promotional machinery," Addresses and Proceedings of the National Education Association, 50:349-55, 1912.

A discussion of changes in school organization that affect promotion.

296. PICKARD, J. L. "Superintendent's report." Twenty-first Annual Report, Department of Public Instruction, Chicago, 1874-75, p. 45-57.

Defends the "graded system" but admits that its evils need correction. Believes that the chief correction should be the promotion of pupils whenever they are ready.

297. PRINCE, J. T. "The grading and promoting of pupils," Educational Review, 15:231-45, March, 1898.

A general discussion with a brief description of several flexible plans.

298. PRINCE, J. T. "Some New England plans and conclusions drawn from a study of grading and promotion," Addresses and Proceedings of the National Education Association, 37:423-34, 1898.

Gives results obtained from a questionnaire, also a description of the Cambridge, Middleboro, Woburn, Keene, St. Louis, Dayton, Le Mars and Centralia plans.

299. PRINCE, J. T. "Classification and promotion of pupils." School Administration. Syracuse: C. W. Bardeen, 1906, p. 125-33.

A discussion of several different flexible plans of organization including frequent promotions, the "double-tillage" plan, the organization of groups according to ability and individual promotion.

300. PYLE, W. H. "A psychological study of bright and dull pupils," Journal of Educational Psychology, 6:151-56, March, 1915.

Shows that mental differences between pupils may be measured.

301. PYRTLE, E. RUTH. "Super-normal children—a study," Journal of Educational Method, 1:273-75, March, 1922.

Describes the plan used in the junior high school of Lincoln, Nebraska. On the basis of their school records the brightest pupils are selected and allowed to do three years of work in two. Many advantages are listed.

302. RATHMANN, C. G. "The Mannheim system of school organization," Educational Review, 53:55-60, January, 1917.

A description of this plan with an account of its effect upon elimination which was much reduced by its operation.

303. RICHMAN, JULIA. "A successful experiment in promoting pupils," Educational Review, 18:23-29, June, 1899.

In a New York school the pupils of grades I-VI were divided into bright, medium and poor sections, each going at its own rate. The effect was to increase the number of promotions.

304. RIGLER, FRANK. "Principles of classification." Thirty-seventh Annual Report of the Public Schools of Portland, Oregon, 1909-10, p. 113-19.

Description of the Portland plan in which the course of study is divided into fifty-four parts arranged in six cycles of one and one-half years each. At the end of each cycle pupils are reclassified according to their ability, there being two rates at which they may progress.

305. ROBERTS, J. E. "A working scheme of promotional efficiency," Elementary School Journal, 17:719-26, June, 1917.

Tells of the plan used in Fond du Lac, Wisconsin, where pupils are promoted at any time by subjects and not by grades.

306. ROSIER, JOSEPH. "The graded school—its strength and its weakness," Review of Education, 7:175-77, December, 1901.

States that the lack of individual instruction is not a weakness inherent but one that can be remedied by teachers.

307. ROSIER, JOSEPH. "A satisfactory basis for the promotion of pupils," School and Society, 1:701-05, May 15, 1915. Addresses and Proceedings of the National Education Association, 53:477-84, 1915.

A theoretical discussion of flexible grading and promotion, arguing for special schools, ungraded rooms, maximum and minimum curricula, etc.

308. REUDIGER, W. C. "Is credit for quality sound?" School Review, 23:450-54, September, 1915.

An argument against giving credit for quality. States that quality and quantity are disparate, that we recognize quality by the honor attached, that credit for quality would over-emphasize marks, and that it would vary the educational content covered by different students but indicated by the same diploma.

309. Sakaki, Yasusaburo. "Some studies on so-called 'ab-normally intelligent' pupils," Psychological Clinic, 6:18-25, March 15, 1912.

Classifies "ab-normally intelligent" pupils into seven classes and shows the need for individual instruction.

310. Search, P. W. "Individual teaching: the Pueblo plan," Educational Review, 7:154-70, February, 1894.

A description of the Pueblo plan in which all work, promotion and graduation is strictly individual. Lists a number of advantages.

311. Search, P. W. "The Pueblo plan of individual teaching," Educational Review, 8:84-85, June, 1894.

A brief explanation of the Pueblo plan supplementary to the reference above.

312. Search, P. W. et al. "Individualism in mass education," Addresses and Proceedings of the National Education Association, 34:398-411, 1895.

A rather general criticism of the graded system followed by discussion of both sides.

313. Shaer, I. "Special classes for bright children in an English elementary school," Journal of Educational Psychology, 4:209-22, April, 1913.

Describes a plan used in a Manchester school in which there are four special rooms by means of which pupils may gain one or more years' time.

314. Shearer, W. J. "The Elizabeth plan of grading," Addresses and Proceedings of the National Education Association, 37:441-48, 1898.

In Elizabeth, New Jersey, promotions were made at any time upon the basis of the teacher's estimate and pupils grouped according to ability, each group going as fast as it was able. Many favorable results were obtained from this plan.

315. Shearer, W. J. The Grading of Schools. New York: H. P. Smith, 1899. 220 p.

Suggests that grading should be in groups according to ability and acquirements.

316. Shearer, W. J. "Grading for efficient organization in the interests of pupils," Addresses and Proceedings of the National Education Association, 40:285-86, 1901.

A general discussion in which the essentials of good grading are listed as follows: acceleration, classification, proper provision for reclassification and proper apportionment of work.

317. Shearer, W. J. "School children in lockstep," World's Work, 14:9252-55, August, 1907.

A strong argument against the lockstep with a discussion of conditions in New York, Philadelphia, etc.

318. SIDERS, W. R. "In class instruction, how can the individual be reached?" Addresses and Proceedings of the National Education Association, 47:175-82, 1909.

Discusses several other plans and describes that in use in Pocatello, Idaho. A combined class, group and individual instruction system is used there.

319. SMITH, H. L. "Plans for saving time in grades VII-XII inclusive." Bulletin of the Extension Division, Indiana University, Vol. 4, No. 4 (Fifth Conference on Educational Measurements), 1918, p. 74-91.

Presents the results of a questionnaire study of the plans in use in over 400 cities.

320. SNEDDEN, DAVID. "Exceptional children." Monroe's Cyclopedia of Education, New York: Macmillan Company, 1911, Vol. 2, p. 540-41.

Says that flexible promotion more or less takes care of exceptional children.

321. SNEDDEN, DAVID. "Grading and promotion." Monroe's Cyclopedia of Education, New York: Macmillan Company, 1912, Vol. 3, p. 126-28.

Mentions several flexible plans and discusses the subject in general.

322. SPAIN, C. L. "The platoon school in Detroit." Detroit Educational Bulletin No. 2, November, 1920. Journal of Educational Research, 3:76-77. January, 1921.

Describes the platoon organization in use in some Detroit schools.

323. SPAULDING, F. E. "The unassigned teacher in the schools," School Review, 15:201-16, March, 1907.

In Newton, Massachusetts, unassigned teachers are assigned to work with individuals and small groups. They help the better pupils to gain time and the poorer ones to keep up with the class.

324. SPAULDING, F. E. "The Newton educational policy." Report of Superintendent. Annual Report of the School Committee in Newton, Vol. 74, p. 15-22, 26, 1913.

States that Newton aims to provide educational opportunities adapted to all individual pupils. Parts of the general plan are to have assistant teachers to help individuals and to make promotions at any time.

325. STERN, WM. "The supernormal child," Journal of Educational Psychology, 2:143-48, 181-90, March and April, 1911.

An argument to the effect that special types of instruction must be provided for the super-normal and sub-normal pupils.

326. STEVENSON, R. W. "Superintendent's report." Columbus Public Schools, 1872-73, p. 45.

States that Columbus, Ohio, allows promotions at any time.

327. TERMAN, L. M. "The mental hygiene of exceptional children," Pedagogical Seminary, 22:529-37, December, 1915.

States that unusually bright children are not given enough work to do and dull children too much.

328. TERMAN, L. M. "Discussion of gifted children section." Proceedings of the First Annual Conference on Educational Research and Guidance, San Jose State Teachers College. Sacramento, California: State Printing Office, 1923, p. 41-43.

A summary of several addresses dealing with gifted children, with some argument in favor of rapid progress rather than enriching the curriculum in the ordinary way.

329. TRUMMER, MARY C. "Instruction adapted to groups of differing ability," School Review, 30:409-11, June, 1922.

Describes the "accomodation" and "express" classes, the "opportunity," "adjustment" and "make-up" groups used in a Los Angeles high school. Pupils are selected on the basis of their elementary school marks.

330. ULRICH, FLORA. "A year's work in a 'superior' class," Psychological Clinic, 5:245-50, January 15, 1912.

In a Cincinnati school a group of superior children did two years of work in one year but this gain was the least of the benefits derived from the plan. Mental activity and development were greatly increased.

331. VAN SICKLE, J. H. "Plan of the North-Side schools of Denver," Addresses and Proceedings of the National Education Association, 37:434-41, 1898.

A description of the Denver plan and a general discussion in favor of flexibility, minimum essentials, etc. In the Denver plan there were no fixed promotion dates until the third grade was reached. Above that the promotions were made every semester.

332. VAN SICKLE, J. H. "Provision for exceptional children in the public schools," Psychological Clinic, 2:102-11, June 15, 1908.

A discussion of provisions for all sorts of exceptional children except the superior.

333. VAN SICKLE, J. H. et al. "Preliminary report of the committee on provision for exceptional children in the public schools," Addresses and Proceedings of the National Education Association, 46:348-85, 47:343-67, 1908, 1909.

A discussion of physically, intellectually, and morally exceptional children and of various provisions being made for them in Worcester, Indianapolis, Baltimore, St. Louis, Berlin, and elsewhere.

334. VAN SICKLE, J. H. "Provision for gifted children in public schools," Addresses and Proceedings of the National Education Association, 48:155-60, 1910.

Mentions various reports and studies that had been made and tells of the Baltimore plan of giving high-school pupils extra work if they are bright enough.

335. VAN SICKLE, J. H. "Provision for gifted children in public schools," Elementary School Teacher, 10:357-66, April, 1910.

States that both inferior and superior children should receive special attention and that there should be groups organized on the basis of ability.

336. VAN SICKLE, J. H., WITMER, L. and AYRES, L. P. "Provision for exceptional children in public schools." United States Bureau of Education Bulletin, 1911, No. 14. 92 p.

A rather complete discussion dealing with methods of finding, classifying, promoting, teaching, and in general providing for exceptional children. Also deals with the number found retarded, eliminated, etc.

337. WALLIN, J. E. W. "The mental health of the school child." Public School Provisions for Mentally Unusual Children. New Haven: Yale University Press, 1914, p. 383-428.

Gives questionnaire results from over 300 cities and sketches the history of special and ungraded classes.

338. WANGER, RUTH. "The special class." News Letter, 19:4-5, May, 1923.

An account of provisions for dull pupils of high-school age. A special course including English, civics, commercial work, sewing, etc. was arranged for those who could not do ordinary academic work. The results have on the whole been favorable.

339. WASHBURNE, C. W. "Breaking the lockstep in our school," School and Society, 8:391-402, October 5, 1918.

An argument for individual progress.

340. WHIPPLE, G. M. "Supernormal children." Monroe's Cyclopedia of Education. New York: Macmillan Company, 1913, Vol. 5, p. 464-68.

A discussion of a number of plans for taking care of bright children.

341. WHITE, E. E. et al. "Several problems in graded-school management," Addresses and Proceedings of the National Education Association, 14:254-63, 1874.

A discussion of annual promotion and the need of adapting the schools to individual pupils.

342. WITMER, LIGHTNER. "The training of very bright pupils," Psychological Clinic, 13:88-96, December 15, 1919.

A discussion of the meaning of brightness, the number of bright and inferior pupils found, etc.

343. WOODROW, HERBERT. Brightness and Dullness in Children. Philadelphia: J. B. Lippincott and Company, 1919. 322 p.

There are various discussions of retardation and advancement, elimination, special classes, bright and dull pupils and experiments with bright pupils.

344. WOODS, ELIZABETH L. "Provision for the gifted child," Educational Administration and Supervision, 3:139-49, March, 1917.

A summary of provisions in a large number of cities showing that flexible promotion is very common, special classes or groups fairly so, and that practically all superintendents favor making some such provision.

345. YOUNG, ELLA FLAGG. "Grading and classification of pupils," Addresses and Proceedings of the National Educational Association, 32:83-86, 1893.

A rather general discussion.

346. "The Cambridge plan," Journal of Education, 75:375-76, April 4, 1912.

A brief description of this plan.

BULLETIN No. 17

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

THE PRESENT STATUS OF WRITTEN EXAMINATIONS AND SUGGESTIONS FOR THEIR IMPROVEMENT

By

WALTER S. MONROE
Director, Bureau of Educational Research

Assisted by

LLOYD B. SOUDERS
Formerly Assistant in Bureau of Educational Research

PRICE 50 CENTS

PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1923

TABLE OF CONTENTS

PREFACE

This bulletin reports the results of three extensive investigations relating to written examinations. These investigations were made by Mr. Souders under the direction of the Director of the Bureau of Educational Research. The tabulations and statistical calculations were made by Mr. Souders or by clerks working under his immediate direction. The preparation of the published report, however, is the work of the Director of the Bureau.

The Bureau of Educational Research wishes to acknowledge its indebtedness to the superintendents, principals, and teachers who cooperated by furnishing the necessary data. The data required in the study of reliability of written examinations necessitated considerable additional labor. Without their cooperation these investigations would not have been possible.

WALTER S. MONROE, *Director.*

November 1, 1923.

PRESENT STATUS OF WRITTEN EXAMINATIONS AND SUGGESTIONS FOR THEIR IMPROVEMENT

CHAPTER I

INTRODUCTION

Preparation and administration of written examinations important phases of the teacher's work. Written examinations, except in the few schools where they have been abolished, form a very important phase of the teacher's work, both because of the time devoted to their preparation and administration and of the significance attached to the measures which they yield. The final grades upon which promotion and the awarding of school honors depend are determined largely by final examinations and by written tests given during the school term. Altho standardized educational tests have become widely used during recent years, written examinations are still the most frequently used type of measuring instrument. This will probably always be true, particularly in the high school. Hence, we may expect that written examinations will occupy in the future as in the past, an important place in the work of our schools.

Need for more information concerning written examinations. There have been numerous investigations which showed that the marking of written examination papers is highly subjective—that is, different teachers tend to assign different marks to the same paper. With the exception of these studies relatively little precise information is available in regard to written examinations but a number of criticisms based upon experience and theoretical considerations have been made. As a result many teachers and other school officials have come to consider written examinations very inferior instruments and have abolished them in a number of schools.

A search through our educational literature, particularly textbooks, reveals an astonishing lack of information in regard to written examinations. Relatively little specific attention has been

given to their preparation and administration in our courses for the training of teachers. Inexperienced teachers have been left largely to their own devices in this important phase of their work. Careful inquiry and observation have indicated that there is a variety of practises with reference to the types of questions asked and to the administration of written examinations. Hence it appears that there is need for a comprehensive investigation of the present status of written examinations in order that a more intelligent estimate may be formed of their value in the process of education and that specific directions may be formulated in regard to their preparation and administration.

Purpose of this bulletin. It is the purpose of this bulletin to present (1) a brief summary of certain previous investigations relating to written examinations and also of the arguments for and against written examinations; (2) the results of three extensive investigations conducted by the Bureau of Educational Research, (a) the preparation and administration of written examinations in Illinois high schools, (b) the constant and variable errors in examination grades, and (c) the content of written examinations; and (3) suggestions for the improvement of written examinations. In the concluding chapter the author presents a list of rules in regard to the preparation and administration of written examinations.

CHAPTER II

SUMMARY OF CRITICISMS OF WRITTEN EXAMINATIONS[1]

Plan of chapter. In this chapter the important criticisms of written examinations are briefly summarized. Following each criticism either a brief answer is given or a reference is made to a detailed discussion in a later chapter. By presenting both sides of the question in this way, it is hoped that the reader will be assisted in forming an intelligent estimate of the merits of written examinations.

I. **Examinations yield inaccurate measures of school achievement.** In support of this argument six points have been made.

1. The most important criticism relating to the accuracy of written examinations is that the marking of the papers is highly subjective. A large number of scientific investigations have yielded objective evidence that different teachers when working independently tend to assign widely varying marks to the same paper. One of the first studies of this type was by Starch and Elliott who found that the marks assigned to the same examination paper in Plane Geometry by 116 teachers ranged from 28 to 92 on the scale of 100 percent. The facts of such investigations as this can not be disputed but as we have no means of securing perfectly accurate measures of achievement, the question at issue concerns the relative rather than the absolute accuracy of the measurements secured. Facts may be misinterpreted. In Chapter IV we shall present evidence to show that when judged in relation to other means for measuring school achievement, written exami-

---

[1]Starch, Daniel, and Elliott, E. C. "Reliability of grading high-school work in mathematics," School Review, 21:254-59, April, 1913.

Morton, Robert L. "The examination method of licensing teachers," Educational Administration and Supervision, 6:421, November, 1920.

Wood, Ben D. "Measurement of college work," Educational Administration and Supervision, 7:301-34, September, 1921.

Kelly, F. J. "Teachers' marking," Teachers College Contributions to Education, No. 66. New York: Teachers College, Columbia University, 1914.

nations yield relatively more accurate measures than generally supposed. In view of the additional information secured the subjectivity of written examinations loses much of its potency as a reason for their abolition.

2. The questions of ordinary examinations are usually not equal in difficulty and weighting by teachers is highly subjective.[2] It has been inferred that this condition tends to increase materially the inaccuracy of examination marks. Comparisons of weighted and non-weighted scores yielded by standardized tests have revealed that the errors introduced by disregarding the unequal difficulty of exercises or questions are not significant in most cases.[3]

3. It has been pointed out that frequently the content of written examinations is not in agreement with recognized educational objectives. Catch questions relating to trivial facts or worded in a misleading way have been cited as illustrations. Certain examination questions also have referred to items which had not been included in the course or at least had received only minor emphasis. Some evidence with reference to the justification of this criticism will be presented in Chapter V.

4. In most examinations the rate of work is neglected. The usual practise is to allow sufficient time for all pupils to finish or to base the mark only on the questions answered in the unfinished papers. Hence a student's examination grade is not influenced by the rate at which he answers the questions. It is easily possible to take into account the student's rate of work in determining the mark assigned to his examination paper. One plan is to set an examination of sufficient length so that all members of the class will be employed during the entire period. Another procedure is to have the student record the time when he finishes. In this way some weight can be given to his rate of work. This criticism is, however, a minor one. In some subjects the rate of work is an important consideration but in others, particularly those in which reasoning predominates in answering the questions, the neglect of the rate of work will affect the accuracy of the examination marks only slightly, if at all.

---

[2]Comin, Robert. "Teachers' estimates of the abilities of pupils," School and Society, 3:67-70, January 8, 1916.

[3]Charters, W. W. "Constructing a language and grammar scale," Journal of Educational Research, 1:249-58, April, 1920.

Monroe, Walter S. "The description of the performances of pupils on exercises of varying difficulty," School and Society, 15:341-43, March, 1922.

5. Written examinations are usually so short that they do not offer an adequate opportunity for a student to demonstrate his ability. This criticism is frequently expressed in the statement that it is unjust to base a student's standing for a semester or a year on an examination paper written during a brief examination period. When stated in this way the criticism refers to two issues between which there is failure to distinguish. The first is in regard to the weight allowed the examination grades in determining a student's final standing. This question of the weight given the final examination is discussed in a later chapter, but it may be said here that the usual practise in high schools is to count the written examination as one-third of the student's total grade. The second refers to the inaccuracy of the grade due to the limited opportunity which is given the student to demonstrate his ability. For practical reasons it is necessary that measurement of the total achievement for the term be based upon a sample. In general, increasing the scope of the examination will tend to increase the accuracy of the measures yielded. Some evidence with reference to the reliability of examination grades based upon short samples will be presented in Chapter IV. It is possible for a teacher to make examinations more comprehensive. This can be accomplished in part by exercising more care in the preparation of the questions. The "new examination" in which pupils are required to do little or no writing affords one means for covering a wide range of subject-matter in a brief period. This method of improving examinations will be discussed in Chapter VI.

The final point to be made with reference to the inaccuracy of examination marks refers to the distinction between a "score" which describes a pupil's performance on the examination and a "grade" which interprets this score with reference to a norm. Failure to recognize this distinction is primarily responsible for too high grading by some teachers and too low by others. Even the same teacher is likely to assign "high grades" on some examinations and "low grades" on others.

In order to understand how norms (standards) are used in connection with the grading of examination papers it is necessary to distinguish between "scores," or measures, and "grades," or marks. A "score" simply describes the performance which has been recorded on the examination paper. For example, a pupil may answer 55 percent of the questions correctly. In this case 55

is his "score." If a certain number of points or credits had been given for each question his score might be 129, or 91, or 217. A "grade" interprets this description with reference to certain norms. A "grade" indirectly describes a pupil's performance on an examination, but it tells also whether the performance is to be considered as above or below passing; whether the pupil is to receive the highest mark or the lowest mark or an average mark.

It is customary to describe the quality of examination papers in terms of the percent of questions answered correctly. For example, if an examination includes ten questions and a pupil answers seven of them correctly and an eighth one partially right, he is given a score of 75 percent, which is interpreted to mean that in the judgment of the examiner he has answered the questions 75 percent correctly. School marks or "grades" are also frequently expressed in terms of percents. Sometimes they are expressed in terms of letters or other symbols, but these in turn are defined in terms of percents. For example, the grade of "A" may be defined as being between 95 percent and 100 percent. Since both "scores" and "grades" are generally expressed in terms of percents, it is only natural that the two have been confused and that "scores" have been used as "grades."

A good illustration of their difference came to the writer recently. An examination in mathematics was given to nearly 1000 freshmen in one of our large universities. This examination may properly be described as "hard," considering the training which the students had received. One student made a score of 100. The lowest score was 12. The average was approximately 55. From the standpoint of the distribution of scores this was a "good examination." If it had been easier, so that any considerable number of pupils received scores of 100 percent, it would have been unsatisfactory. If it had been so "hard" that a considerable number of students made zero-scores it would also have been defective. In both cases it would have failed to differentiate between some students who were not equal in ability. However, obviously an injustice would be done if a passing mark of 70 or 75 were adopted and all pupils having scores below this mark were given a grade of failure. The passing mark for this particular examination should be in the neighborhood of 40. If the "scores" are to be represented in terms of "grades" a "score" of 40 should be translated into a "grade" of 70 or whatever passing mark has been adopted by the institution.

The recognition of this distinction between "scores" and "grades" enables us to indicate the way in which subjective norms are implied in "grades." A "grade" is not a pure measure or description of the pupil's performance. It is rather an interpretation of the measure of his performance with reference to certain norms. When no distinction is made and "scores" are used as "grades," pupils will receive high "grades" if the examination is "easy;" and low ones if it is "hard." Thus, the difficulty of the examination is one factor in establishing the norms with reference to which the "scores" are interpreted when they are used as "grades." Severe marking will tend to set high norms. Only when the examination is of average or "standard" difficulty and the marking is average in severity do "scores" and "grades" become identical in magnitude. Since the norms are established by the difficulty of the examination and the severity of the scoring, they must be subjective. In the investigations of the marking of examination papers it was shown that teachers varied widely in their judgments concerning the worth of examination papers. There is no reason to expect that they would agree more closely in estimating the difficulty of examinations. Hence, norms which depend upon teachers' estimates of the questions appropriate for examinations and upon their marking of the papers must be considered subjective. It is possible to increase greatly the objectivity of these norms and the first requirement is to recognize the distinction between "scores" and "grades." (See page 38 for a further consideration of this topic.)

Summary of inaccuracy of examination marks. From the preceding discussion examination marks are, without doubt, shown to be far from accurate measures of school achievement. However, it does not necessarily follow that the errors involved are of sufficient magnitude to justify the abolishment of written examinations. In the writer's belief the greatest benefit will come from making an intelligent inquiry into the nature of these errors and from taking steps to reduce them to the lowest magnitude.

II. Written examinations tend to encourage undesirable mental processes. Many critics have claimed that most examinations, particularly those given at the end of a course, tend to encourage "cramming." The assertion is made that many students do little or no studying until near the close of the term. Then by the process of "cramming" they are able to pass the final ex-

amination and attain a relatively high standing in the course. This criticism assumes that "cramming" is an undesirable mental process and that final examinations are responsible for its occurrence. The undesirable feature is the neglect of study throughout the term. This is not due to the fact that final examinations are given but that undue emphasis is placed upon them and that the teacher has failed to check up on the student's work day by day throughout the term.

One of the points which may be made in favor of final examinations is that they furnish an immediate incentive for review and organization of the content of the course. The writing of an examination itself may be an important part of the student's learning. This is particularly true in the case of questions which require reasoning and organization of information. "There is no impression without expression," and the writing of a three-hour examination is undoubtedly an intensive form of expression. Hence, one is justified in maintaining that written examinations tend more to encourage desirable mental processes than undesirable ones.

III. Passing the final examination an undesirable objective. The assertion has been made that when a final examination is required, the passing of it tends to become the objective for which many students work. When this occurs it is due not to the fact that the final examination is required but rather to the undue emphasis which is placed upon it by the school. If an examination consists of appropriate questions it is not undesirable to have the student keep it in mind as one of the objectives to be attained by studying the subject-matter of the course. However, as we shall show later, (see page 25) the usual practise is to count the final examination grade as one-third in determining a student's final standing. In many schools it receives less weight. When the final examination counts only one-third or less in determining a student's final standing it is difficult to say in what respect it forms an important educational objective.

IV. Examinations injurious to health of students. Some critics claim that written examinations, particularly those given at the end of a course, are injurious to the health of students, many of whom make very strenuous preparation for them. The obvious strain which accompanies the writing of answers to the questions of examinations sometimes lasting two or three hours

must also be borne. It is undoubtedly true that both the preparation and the writing frequently make enormous drains on the energies of students. However, no careful investigation has been conducted of the actual effect upon their health. To one who observes the great expenditures of time and energy devoted to social and athletic activities, it is difficult to believe that examinations are in general more injurious to the health of students than many other activities in which they are permitted and even encouraged to engage. Here again it should be realized that this criticism is not fundamentally a criticism of examinations, but rather of setting very long examinations or of placing extreme emphasis upon them by making the final grade of the course depend wholly or very largely upon the examination grade.

V. Time devoted by teachers to written examinations not profitably spent. In the opinion of some critics the time given to the preparation of questions and particularly to the marking of examination papers might be more profitably employed. Information concerning the time actually devoted to the preparation and the administration of written examinations is given in Chapter III. However, it may be pointed out here that a teacher can not attain a high degree of efficiency as an instructor unless he checks up the work of his students in order to assist those who need supplementary and remedial instruction. Only by knowing the extent to which his students have achieved individually and collectively can a teacher make his instruction fit the needs of his class. Thus considerable time must be given to measuring the results of teaching. This is an indispensable portion of the teacher's task. It is only when a teacher devotes an undue proportion of his time to the preparation and administration of examinations that such work tends to be wasted. Doubtless, the time devoted to written examinations might in many cases be profitably increased. Students receiving low marks should have their answers studied in order to ascertain in what ways and why they have failed. Such information will frequently be exceedingly illuminating to the instructor, and aid him in determining his own shortcomings.

CHAPTER III

PREPARATION AND ADMINISTRATION OF EXAMINATIONS IN HIGH SCHOOLS

The data collected. The purpose of the study reported in this chapter was to secure information concerning the present practise in the preparation and administration of written examinations in high schools. A questionnaire was mailed in the fall of 1922 to 254 high-school principals in Illinois and a second one was sent to approximately 2900 high-school teachers.[1] One hundred and eighty-nine replies were received from principals and 1816 from teachers. Of the latter it was necessary to discard eighty so that the following report is based upon returns from only 1736 high-school teachers who are distributed as follows:

Commercial Subjects	192	Modern Languages	82
Drawing and Art	26	Music	21
English	342	Science	309
Home Economics	143	Shop Work	58
Latin	118	Social Science	198
Mathematics	247		

Representative character of data collected. The high schools from which answers to the questionnaire were received ranged from those established in rural communities to a large metropolitan high school. No supplementary investigation was made to ascertain the extent to which the replies were representative of conditions in Illinois but in the tabulations there was no indication that the data collected were not representative of the state. A few of the replies, particularly those of teachers, suggest that some slight misinterpretation of certain of the questions may have been made. (See page 23). Such cases, however, were relatively rare and probably did not affect the median of the results.

Extent of the requirement of final examinations in Illinois high schools. Evidence of the subjectivity of the marking of examination papers, together with other adverse criticisms of written examinations, has tended to cause many teachers and superintendents to be skeptical of their value. In a number of schools

[1]These questionnaires are reproduced in the appendix on pages 66 and 68.

final examinations have been abolished or made optional with the teachers and they are not considered essential by many teachers. In order to ascertain the present practise in Illinois the high-school principals were asked, "Do you require your teachers to give final examinations?" Only twenty-one principals or 11 percent stated that final examinations were not required. Thus it is the practise in Illinois high schools to require that final examinations be given. This, however, does not mean that all students must take them. Of the 168 high schools in which final examinations are required 101 or 60 percent reported that it was their practise to exempt certain students. Scholarship, that is making a grade on daily work above a certain average, was mentioned by all of these schools as one of the conditions on which exemption was based. Deportment was mentioned by 52 percent and attendance by 32 percent as additional conditions.

No information was secured with reference to the explanation of the exemption from examinations of students meeting certain conditions but general observation has indicated that two reasons are frequently recognized. The first is that promise of exemption from the final examinations operates as a powerful motive to secure a high quality of daily work, regular attendance, and good deportment. The other is the belief held by many teachers that final examinations are unnecessary to determine a student's standing in a course. They contend that the average of a student's daily grades should be taken as a final grade for the course.

There is no doubt that the promise of exemption from the final examination operates as a powerful motive in the case of many students. It should, however, be recognized that such an incentive is artificial and therefore open to criticism. In so far as possible a student should be actuated by motives which sustain an intrinsic relation to the subject-matter. If it is necessary or advisable that the final examination be considered as a motive, it could be used to encourage systematic review and organization of the course. This should constitute a very important phase of studying. Students may, of course, be asked by their teachers to review frequently and to summarize and organize the work at the end of the term, but they cannot be convinced easily of the necessity of such work if it receives no more weight in determining their final grade than their performances during an equal period of time elsewhere in the course.

The second reason is a valid one in many cases. In the experience of most teachers the mark made on the final examination changes the standing of relatively few students. Experienced teachers can under favorable conditions estimate with considerable accuracy the achievements of their pupils. If the class is reasonably small and if the teacher has used methods of instruction which call for frequent oral and written performances by the students and has kept a careful record of these performances throughout the term, his estimates will generally be relatively accurate measures of the achievements of the students. There are, however, certain limitations which should be noted. Teachers may be unduly influenced in their estimates by the more recent performances of their students. Unless careful records have been kept throughout the term inferior work at the beginning tends to be overshadowed by good or excellent work during the closing weeks. In case the class is a large one the teacher may not have an adequate opportunity for becoming acquainted with all of its members.

Teachers' estimates are likely to be materially affected by personal characteristics of students; one with a pleasing personality is in many cases rated higher than one who is unattractive. If the classwork is conducted so that there is little or no written performance required, teachers' estimates will necessarily be based almost wholly on the oral responses given during the class period. Some pupils make a good showing in class when the recitation is oral but are at a decided disadvantage when asked to record their answers in writing. Frequently this difficulty is encountered when they are careless in their thinking and do not have clear ideas to express. In oral recitation they are able to make a fair showing because of personal characteristics and because of the stimulus of detailed questioning by the instructor. Furthermore, in a class discussion a bright student who has a good command of language may easily pick up ideas from other members of the class and recall ideas from his general experience sufficient to make a good showing. On the other hand there are students who express themselves more effectively in writing. They may be good thinkers but a little slow in their mental processes and not clever in discussion. Thus there are cases in which it is difficult or impossible for a teacher to estimate accurately the real achievements of students from their daily work alone. The final examination at the

end of the term will in a considerable number of cases furnish additional information which is needed in arriving at the student's true standing.

The final examination in itself provides a distinct type of educational opportunity which does not occur elsewhere in the course. Altho the writers have no evidence to present upon this point they are convinced from their experience with college students and from the comments of a number who have been exempted from final examinations in high school that the practise deprives students of an important educational opportunity. Not infrequently students who have been "excused from examinations" in high school state that they experienced a distinct handicap when they entered college. If final examinations can be justified they should be required of all students. To use them only as a device for motivating the work of the term destroys much of their value.

Time devoted to written examinations. Three questions were asked relative to the time devoted to the preparation and administration of written examinations. The replies from the principals indicated that the most frequent practise is to allow ninety minutes for the writing of a final examination. This is the time allowed in 45 percent of the schools having final examinations. Fifteen percent allow eighty minutes and a slightly larger percent one hundred and twenty minutes.

The teachers were asked to state approximately the number of minutes which they use "in preparing questions for a final examination which students are allowed a total of ninety minutes to answer." The median time which varies only slightly for the different subjects is approximately fifty minutes. Individual teachers in the same subject differ widely in the amount of time which they give to this phase of their work. Two teachers, one in mathematics and one in science, stated that they spent more than six hours in the preparation of a set of final examination questions. In each subject there were a number of other teachers who stated that they devoted not more than thirty minutes to such work. It is possible that some teachers failed to interpret this question correctly but doubtless much of the variation is due to differences in the practises of teachers during the semester. Some probably make a memorandum of questions as they occur during the term and use this list as a basis for preparing the final examination. Also experience is a contributing factor. Teachers who have be-

come very familiar with the subject should be able to formulate questions more quickly than those who are not so well versed.

The teachers were asked also to give the approximate time which they used in "marking the papers of a final examination which students are allowed a total of ninety minutes for answering." They were directed to base their estimates upon a class of twenty-five students. The median time is approximately two and one-half hours. The variations between the different subjects are not large when the differences in their character are considered. The greatest number of hours are required for English and social science and the least for drawing and art. Here also there were wide variations in the amount of time reported by the individual teachers. A total of twenty-five teachers in which all subjects except home economics, shop work, and social science were included, stated that they devoted not more than thirty minutes to marking a set of papers for twenty-five students. On the other hand, thirty-nine teachers stated that they spent 480 minutes or eight hours in the marking of a single set of papers.

It is obvious from the replies received that some teachers treat this phase of their work much more seriously than others or that they employ widely different methods. Probably some correct all errors or insert references which will enable the students to correct their own errors when the papers are returned. Others merely check the errors and still others probably do not attempt to even check each error but estimate the worth of the paper as a whole. The question concerning the amount of time which a teacher is justified in devoting to the marking of a set of examination papers may very profitably be raised. Final examination papers should be treated seriously and there should be an earnest endeavor on the part of the teacher to estimate as accurately as possible the grades which are assigned but it is doubtful if the expenditure of as much as twenty minutes per paper which was reported in some cases could be justified. The median practise seems to represent a more reasonable amount of time.

Characteristics noted in marking examination papers. The principals were asked to state whether it was the practise in their schools for teachers to subtract from a pupil's grade for (1) poor writing, (2) poor spelling, (3) poor English. Seventy-one principals or 42 percent stated that teachers were accustomed to make deductions for poor writing. In 60 percent of the schools it was the

TABLE I. PERCENT OF TEACHERS REPORTING INTENTIONALLY LOWERING A STUDENT'S GRADE FOR POOR WRITING, POOR SPELLING, AND POOR ENGLISH

Subject	Poor Writing	Poor Spelling	Poor English
Ancient Language	31	75	80
Commercial Subjects	61	74	72
Drawing and Art	65	82	67
English	60	96	98
Home Economics	37	72	73
Mathematics	31	48	48
Modern Language	31	82	74
Music	48	60	48
Science	35	53	56
Shop and Vocational	32	77	74
Social Science	37	61	65

practise to lower a student's grade for misspelled words and in 68 percent for poor English. Fifteen principals or 9 percent reported that all three characteristics were recognized only in the marking of papers in courses in English.

The teachers were asked if they intentionally lowered a student's grade because of each of the three characteristics mentioned in the above paragraph. A summary of their replies is given in Table I, which indicates considerable variation with reference to the influence of poor writing, poor spelling, and poor English upon examination grades. Since writing, spelling and English may be considered essential parts of courses in English we should naturally expect that teachers of this subject would intentionally lower a student's grade for defects in any of these characteristics. Outside of the subject of English, the majority of teachers do not lower a grade for poor writing except in commercial subjects, and drawing and art. With the exception of mathematics deduction is made by most teachers for poor spelling. The potency of poor English in determining a student's grade is slightly less than that of spelling in a number of subjects.

The handwriting, spelling, and quality of English which a student uses in writing an examination should be recognized. It does not, however, follow that a student's standing should be intentionally lowered for poor handwriting, poor spelling, and poor English in school subjects other than English. When this is done his grade becomes a measure of these abilities as well as of the abilities in the field of the subject in which the examination is given. In history, for example, a student's grade would become a

composite measure of his achievement in history, the legibility of his handwriting, the quality of his spelling and the use of grammatically correct English. As a result both the teacher and the student are likely to be confused concerning the shortcomings of the examination paper. A better procedure would be to keep a record of the errors in spelling, poor writing, and poor English and when it is considered desirable a separate grade may be given covering these three characteristics. Credit for a course may be withheld until the student has brought his handwriting, spelling and English up to a satisfactory standing.

The weighting of questions. Sixty-four percent of the high school principals indicated that their teachers were accustomed to give more credit for correct answers to difficult questions than to easy ones. Approximately four-fifths of the teachers replying to the questionnaire stated that they attempted to weight examination questions on the basis of difficulty. Thus there is a very definite effort to eliminate the errors introduced in examination grades by the unequal difficulty of questions. (See page 10.)

Recognition of rate of work. Eighty-two percent of the teachers stated that they were accustomed to set examinations short enough so that practically all students could answer all the questions. Only 32 percent noted the time which each student spent in writing his examination paper, and only 8 percent said it was their custom to set examinations long enough so that practically no student would have time to answer all of the questions. Thus it is clear that relatively few teachers recognize the student's rate of work in determining his grade on an examination.

Incidentally it may be noted that when examinations are short enough so that practically all students can finish a great deal of time is wasted. Individual differences exist in all classes and it is not at all unusual to find some student finishing in one-third to one-half of the time which others devote to the examination. Aside from the waste of time which results from this practise, it is likely that the confusion caused by the leaving of those pupils who have finished tends to disturb the attention of those who are still writing. If final examinations constitute a valuable educational opportunity, there is no justification for wasting time. It is much better to set an examination long enough so that practically all students will be occupied for the entire period.

TABLE II. PERCENT OF TEACHERS GIVING AFFIRMATIVE ANSWERS TO FOUR QUESTIONS RELATING TO THE MARKING OF EXAMINATION PAPERS

Subject	Correct Answers Written	Each Ques. on one paper	One Ques. on all papers	Each paper as a whole
Ancient Language	9	73	24	20
Commercial Subjects	48	82	16	31
Drawing and Art	40	75	25	42
English	22	76	19	34
Home Economics	19	68	24	38
Mathematics	72	75	18	24
Modern Language	15	74	23	33
Music	35	80	21	42
Science	27	72	27	27
Shop and Vocational	36	75	27	37
Social Science	18	75	22	40

Method of marking examination papers. Scientific investigation has revealed that the reliability of examination grades can be materially increased by adopting a systematic method in marking papers.[2] Among the procedures recommended are the writing out of correct answers, and the grading of one question on all of the papers before taking up another question. In order to ascertain the practise of high-school teachers in marking examination papers the following four questions were included in the questionnaire sent to them.

1. Before starting to grade a set of examination papers do you write out the answers which you consider correct?
2. Do you usually grade all the answers on one paper before taking up those of another paper?
3. Do you usually grade the answers to one question on all of the papers before taking up the answers to a second question?
4. Instead of marking the answers to each question separately do you attempt to estimate the value of the paper as a whole?

The percent of teachers giving affirmative answers to these questions is given in Table II. A few apparent discrepancies in this table are due to the fact that certain teachers did not answer all the questions. Normally we should expect a teacher who answered the second question affirmatively to answer the third one negatively. This, however, did not always happen.

[2]Kelly, F. J. "Teachers' marks," Teachers College Contributions to Education, No. 66. New York: Teachers College, 1914, 83 p.

With the exception of mathematics, it is not the custom of teachers to write out the answers to their questions. No data are at hand to show what effect this has upon the accuracy of the examination marks. Experience with standardized tests would indicate that the failure to write out correct answers, at least in abbreviated form, would operate to make the grading of examination papers less accurate.

About three-fourths of the teachers are accustomed to mark all the questions on one paper before taking up another. This plan has the advantage of enabling the teacher to consider a pupil's performance as a whole. In the case of students who make a large number of errors the teacher will find this helpful in providing remedial instruction. It has, however, been proposed that the reliability of examination grades can be increased by marking the answers to one question on all the papers before taking up the answers to another question.

Directions to students concerning methods of work. The high-school teachers were asked the following question, "Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answer, arrangement of work, etc.)" Only 35 percent of the teachers gave an affirmative answer. It is possible that in some classes there is a sufficiently definite understanding concerning the methods of work to be followed and explicit directions are unnecessary. However, it is likely that in many cases students would be able to give more truthful evidence of their ability if they were given precise directions concerning the length of answer, desired arrangement of work, etc. The order in which the questions are to be answered is a point which should be stressed. In the case of questions that are at all indefinite or general there should be specifications concerning the degree of elaborateness which is expected in the answer.

Recognition of a standard distribution in assigning grades to examination papers. The teachers were asked the following question, "In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?" Only 31 percent of the teachers gave an affirmative answer to this question. This probably means that

relatively few teachers have recognized the distinction between "scores" and "grades," (See pages 11-13 for an explanation of this distinction.) and for this reason are neglecting one means of making their grades more accurate measures of school achievement.

Relation of examination grades to final grades. The principals were asked if they advised their teachers as to the proportion of the final mark for the semester which should be based upon the final examination. Only eleven or 7 percent replied negatively. Of those who gave advice on this matter 95 percent made a definite ruling. The most frequently mentioned proportions allotted the final examination are 25, 30, 33⅓, and 40 percent. In 4 percent of the schools the examination counts for one-half in determining a student's final grade, in 1.3 percent for only one-tenth of the final grade. The teachers were also asked this same question. The replies varied from 10 percent to 50 percent. Except in science and shop work, the median practise is to estimate the final examination mark as one-third in determining a pupil's final grade. A considerable number of teachers indicated that they gave not more than one-fourth or one-fifth value to the final examination.

Summary. The typical practise with reference to final examinations in Illinois high schools may be summarized as follows:

1. Final examinations are required of students and exemptions are made largely on the basis of scholarship.

2. Students are allowed ninety minutes for writing a final examination. Teachers spend slightly less than one hour in preparing examination questions and from two to to three hours in grading a set of papers for twenty-five students.

3. With the exception of mathematics, the majority of teachers lower a student's grade for spelling, and with the exception of mathematics and music the majority lower it for poor English. Poor writing is not a potent factor except in English, commercial subjects, and drawing and art.

4. About three-fourths of the teachers attempt to weight the questions on the basis of difficulty.

5. The majority of the teachers do not consider rate of work in estimating the grade assigned the final examination paper.

6. The majority of teachers do not write out the answers to the questions preparatory to the marking of papers. The general practise is to mark all answers on one paper before taking up the next.

7. About one teacher in three writes out directions with reference to the procedure which the students shall use in answering the questions.

8. About one teacher in three tries to have his grades conform to a standard distribution.

9. The proportion of the final mark for the semester which is based upon the final examination grade varies from 10 to 50 percent. The median practise is $33\frac{1}{3}$ percent. The majority of the principals make a definite ruling regarding the value placed upon the final examination.

CHAPTER IV

THE CONSTANT AND VARIABLE ERRORS IN EXAMINATION GRADES

Constant and variable errors of measurement.[1] Two types of errors are encountered in educational measurement. The presence of *variable errors* is indicated when a test is given twice to the same group of pupils. The two average scores of the group may be the same but this will not be true for individual pupils. A few pupils will make the same or approximately the same score on the two trials. Others will make higher scores on the second trial than on the first, while still others will make lower scores on the second trial than on the first. If we assume that the average of the two scores obtained represents an approximately true measure of a pupil's achievement then the differences between the first set of scores and the corresponding average scores would be the variable errors of the measures resulting from the first application of the test. Some of these differences approximate zero. Some of them are positive and about an equal number negative. Another set of variable errors would be obtained by using the scores secured from the second application of the test. In the case of a number of pupils the variable errors for the first application of the test will not be the same as those for the second application. Thus, as the name implies, variable errors change in magnitude from pupil to pupil within a group and also for the same pupil in a series of measurements of the same achievements.

A *constant error* is the same for all members of a group. Such an error occurs in teachers' marks where there is a tendency to grade too high or too low. It is found in the case of standardized educational tests when mistakes occur in the time allowed or when other departures are made from standard testing conditions. A constant error may be either positive or negative and it is generally different for different tests.

[1]For a more detailed discussion of the nature and magnitude of the constant and variable errors of educational measurement, see Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923. 30 p.

These two types of errors usually occur in combination, that is, a given measurement may and frequently does involve both a constant error and a variable error. The actual error is a combination of these two. However, in studying the accuracy of educational measurementsit is helpful to distinguish between the two types and to consider each separately. The usual method used for calculating an index of the magnitude of the variable errors does not give any indication of the magnitude of the constant error. Also the method commonly used for determining the presence and probable magnitude of constant errors does not yield an index of the variable errors. Furthermore, different methods are required for decreasing the two types of errors in educational measurements.

Methods of describing the magnitude of the variable errors of measurement yielded by standardized educational tests. In describing the magnitude of variable errors in the measures yielded by standardized educational tests, the usual method is to have the test given twice to a typical group of pupils under as nearly the same conditions as possible. The coefficient of correlation between the two sets of measures is taken as the index of the magnitude of the variable errors. Usually there will be a constant error in one and sometimes in both of the sets of measures but the statistical procedure used is such that this error does not affect in any way the coefficient of correlation secured. This coefficient of correlation is commonly spoken of as the coefficient of reliability. A coefficient of 1.00 would mean that the variable errors were zero.

Available data with reference to the magnitude of errors of examination grades and standardized test scores not comparable. Investigations of the Starch-Elliott type have proven that examination grades involve errors but the method which they employed is different from that used in studying the errors in the scores yielded by standardized educational tests. Starch and Elliott confined their efforts to a study of the subjectivity of the marking of a single examination paper. Except in the use of quality scales, as in handwriting and English Composition, the scoring of standardized educational tests has been made highly objective. Hence, there has been little need for studying the subjectivity of the marking of test papers. On the other hand, there

has not been, so far as the writers are aware, any reported attempt to apply to written examinations the method commonly used in studying the reliability of standardized educational tests. Hence, comparisons to show the relative reliability of the two types of measuring instruments cannot as yet be made.

For this reason in the present investigation it has seemed worth while to apply to written examinations the same method which is commonly used in studying the reliability of educational tests. Certain modifications are of course necessary. These will be noted in the following paragraphs. The investigation pertains primarily to the variable errors involved in examination grades. Incidentally some light will be thrown upon the magnitude of constant errors.

Methods employed in the present investigation of the reliability of written examinations. The essential feature of the methods employed in the present investigation is securing two independent examination grades for each pupil for the same units of work. This requires that two examinations be given to each of the groups of students from which data were secured. Two methods were used. These are described in the following directions which were sent to those cooperating in this investigation.

Method I

Two sets of examination questions are to be prepared by a single person, or two or more persons working together. Each of the two lists should contain the same number of questions. There should be a distinct effort to make the two lists approximately equal in difficulty and as nearly as possible similar in respect to the type of questions.

After the two lists of questions have been made both should be given by each teacher to all of her pupils under as nearly the same conditions as possible. If not given on the same day, the two examinations should be given within a period of one week. For example, if two sets of examination questions in seventh grade geography have been prepared, both sets of questions should be given by each seventh grade teacher to all of her pupils.

Each teacher is to mark both sets of examination papers for her pupils. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination paper. When two or more teachers have given the same examinations it is not necessary that they confer in regard to the marking of the papers. If this is done a memorandum regarding the procedure should be attached to the examination papers.

This method may be followed by a single teacher who has two or more sections of a given subject. Both examinations should be given to all sections taught by this teacher. This method of studying the reliability of written examinations can be applied to any school subject. The Bureau of Educational Research is most interested in having it applied to arithmetic, history, geography, and language in the elementary school and to history, mathematics, English, and science in the high school.

Method II

Two sets of examination questions for the same subject are to be prepared by two teachers working independently, each teacher preparing a set. There is no requirement concerning the length or the difficulty of the two sets of examination questions except that both should cover the same amount of work. The teachers who prepare the questions should not confer concerning either the kind or the number of questions asked.

After the questions have been prepared, both sets are to be given by each teacher to all of her pupils. If not given on the same day, the two examinations should be given within a period of one week.

After the examinations have been given each teacher will grade all of the papers written upon the questions that she prepared. This will mean that she will grade a set of papers for her own pupils and also a set for the pupils of the other teacher. There should be no conferring between the teachers in regard to the method of scoring. In marking these papers the teacher should indicate the credit given for each question and write the total grade plainly upon the examination papers.

The data collected. Through the city superintendents and high-school principals a general invitation was extended to school systems in Illinois to participate in this investigation at the close of the second semester 1921-22 and also at the close of the first semester 1922-23. No instructions other than those just noted were given to those who cooperated. It should, therefore, be borne in mind that the data collected are for written examinations as they are usually given and not for special types of examinations or for unusual methods in the administration or the grading of the test papers. The reliability of examination grades could probably have been increased if certain directions had been formulated in regard to the marking of the examination papers but the purpose of the investigation was to determine the reliability of typical written examinations administered in the usual way.

Returns were secured from seventy-two groups of children but it was necessary to discard the data for six groups because

instructions had not been followed. The examinations given to the sixty-six groups were all of the traditional type. The papers were marked by the teachers on the scale of 100 percent, and were then sent to the Bureau of Educational Research. The coefficients of reliability reported in this chapter were calculated under the direction of the writers.

Coefficients of reliability of examination grades. The coefficients of reliability of written examinations for the sixty-six groups of students are summarized in Table III. This table also shows the number of students in each class, the number of questions in each examination and the method followed in giving the examination. The reliability coefficients have been grouped by subjects and have been arranged in descending order of magnitude. For those entries marked with an asterisk (*) in the column headed "Method," one of the examinations was given by the principal or some other person not actually teaching a class in the subject at that time. However, this person was considered competent to be in charge of the examination. The total distribution of reliability coefficients is given in Table IV.

Two of the coefficients of reliability are negative. The highest is .95. It is interesting to note that the coefficients given for history are, on the average, higher than those obtained for arithmetic. The most reliable examinations given were in algebra. With the exception of history, arithmetic, and algebra, the number of groups is so small that comparisons can not have much significance. The median coefficient of reliability .65 may be used as a general index of the reliability of written examinations.

The coefficients of reliability of standardized educational tests. McCall[2] has stated that the "range of self-correlation for many standardized tests is about .5 to about .9." The writer's experience has indicated a somewhat greater range. In Table V the reliability of a number of standardized educational tests is given. Those for the silent reading tests by Brown, Starch and Courtis are taken from a recent bulletin[3] by the writer. The range in this table is from .19 to .92.

---

[2]McCall, W. A. How to Measure in Education. New York: The Macmillan Company, 1922, p. 396.

[3]Monroe, Walter S. "A critical study of certain silent reading tests." University of Illinois Bulletin, Vol. 19, No. 22, Bureau of Educational Research Bulletin No. 8. Urbana: University of Illinois, 1922, p.33-34.

TABLE III. COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
64 (Arithmetic)...........	21	10	.76	II
58 "	35	7 and 8	.74	II
43 "	24	10	.73	II
63 "	38	10	.71	II
61 "	64	10	.69	II
2 "	37	5	.67	I
62 "	38	10	.64	I
1 "	88	10	.64	I
60 "	41	5 and 10	.61	II
65 "	33	10	.60	II
4 "	72	6	.56	I
71 "	17	10	.48	*
44 "	22	7 and 10	.48	II
3 "	55	6	.47	I
66 "	73	10	.47	II
23 "	27	5	.33	I
32 "	56	5	.30	II
59 "	21	6 and 10	.29	I
45 "	43	10	.06	II
42 "	27	8 and 10	—.18	II
5 (Algebra)	51	10	.92	I
72 "	20	30 and 19	.91	*
50 "	23	5 and 7	.88	II
34 "	37	10 and 6	.82	II
57 "	36	10	.81	II
6 "	52	6	.78	—
35 "	45	15	.73	II
31 "	54	6	.61	II
7 (Language)	33	5	.68	—
9 "	54	5	.62	—
8 "	53	5	.57	—
10 (Literature)	74	5	.19	I
18 (English)	15	3	.78	I
12 "	43	5	.53	II
11 "	27	5	.53	I
54 "	37	7 and 45	.50	II
53 "	21	6 and 7	.47	II
13 "	52	5 and 8	.41	II

TABLE III. (CONTINUED) COEFFICIENT OF RELIABILITY FOR WRITTEN EXAMINATIONS SET BY TEACHERS

Class Number	No. of Pupils	No. of Questions	Coefficient Correlation	Method
20 (History)	14	5	.95	—
15 "	28	5	.85	I
23 "	19	5	.78	I
36 "	43	5	.76	II
56 "	19	9	.75	*
67 "	53	5	.75	II
41 "	32	8 and 10	.67	II
40 "	28	5 and 7	.66	II
14 "	64	5	.63	I
48 "	24	10	.57	II
49 "	30	5 and 10	.55	II
39 (Geography)...........	29	10	.66	II
46 "	29	5 and 10	.66	II
47 "	47	5 and 10	.62	II
16 "	21	5	.43	I
68 "	23	10 and 8	.32	*
38 "	26	10 and 5	—.11	II
17 (Civics)	63	5	.39	I
37 (Latin)	31	8	.89	II
19 "	30	2	.82	I
51 "	50	6 and 8	.68	II
21 (Spelling)	33	5	.87	I
52 (Geometry)...........	42	5	.75	II
70 "	10	8	.68	*
22 "	46	4 and 5	.34	—
24 (Spanish)	18	5	.79	I
33 (German)	23	7 and 8	.83	I
69 (Commerce)...........	16	16 and 10	.53	*

TABLE IV. SUMMARY DISTRIBUTION OF COEFFICIENTS OF RELIABILITY FOR WRITTEN EXAMINATIONS

Size of Coefficient of Correlation	Frequency
.95	1
.90	2
.85	4
.80	4
.75	9
.70	4
.65	9
.60	8
.55	4
.50	4
.45	5
.40	2
.35	1
.30	4
.25	1
.20	0
.15	1
.10	0
.5	1
0	0
— .5	0
—.10	0
—.15	1
—.20	1
Total.........	.66
Median.......	.65

From certain unpublished studies by the writer the following information has been obtained. The Courtis Standard Research Test, Series B, Forms 1 and 2 were given to pupils as follows: Grade V, 89; Grade VI, 81; Grade VII, 52; and Grade VIII, 38. The thirty-two coefficients of reliability ranged from .409 to .904 with the median at .665. Forms 1 and 3 were given to a slightly larger group in each of the four grades. The thirty-two coefficients of correlation between the two sets of scores for this administration of Series B ranged from .528 to .963 with the median at .704. The Woody Arithmetic Scales, Series A, were given to several groups of pupils. Two scores were secured by using alternate items of each of the scales and applying Brown's formula.[4] The twelve coefficients of reliability computed in this way ranged from .91 to .46 with the average at .66. Forms 1 and

[4] $r_{12} = \frac{2r_h}{1+r_h}$ In this formula r_h is the correlation between two scores which this test yields. One is based upon reproduction and the other upon answers to questions.

TABLE V. RELIABILITY COEFFICIENTS OF STANDARDIZED EDUCATIONAL TESTS

Test	Coefficient
Illinois General Intelligence Scale*	.92
Courtis Standard Research Tests, Series B‡	.87
Brown Silent Reading Test—Rate	.86
Courtis Silent Reading Test, No. 2—Rate	.85
Otis Group Intelligence Scale§	.84
Monroe Standardized Silent Reading Test Revised*—Rate	.84
Courtis Silent Reading Test, No. 2—Comprehension—No. Quest	.80
Starch Silent Reading Test—Comprehension—Words	.77
Monroe General Survey Scale in Arithmetic*	.76
Monroe Standardized Silent Reading Test Revised*—Comprehension	.76
Monroe Standardized Silent Reading Test Revised*—Rate	.75
Monroe Standardized Silent Reading Test Revised*—Comprehension	.72
Starch Silent Reading Test—Comprehension—Ideas	.72
Indiana Attainment Scale, No. 1‡	.66
Starch Silent Reading Test—Rate	.62
Pressey Primer Scale†	.59
Courtis Silent Reading Test, No. 2—Comprehension—Index	.58
Pressey First Grade Vocabulary Scale‡	.37
Brown Silent Reading Test—Comprehension—Quantity	.36
Pressey Primer Scale†	.33
Brown Silent Reading Test—Comprehension—Quality	.19

*Monroe, Walter S. "The Illinois Examination." University of Illinois Bulletin, Vol. 19, No. 9, Bureau of Educational Research Bulletin No. 6. Urbana: University of Illinois, 1921, p. 47.

†Pressey, L. W. "A Group Scale of Intelligence for Use in the First Three Grades: its validity and reliability," Journal of Educational Research, 1:285-94, April, 1920.

††Unpublished data of the Bureau of Educational Research, University of Illinois.

§Colvin, S. S. "Some recent results obtained from the Otis Group Intelligence Scale," Journal of Educational Research, 3:1-12, January, 1921.

2 of Monroe's Standardized Reasoning Test in Arithmetic were given to pupils as follows: Grade V, 36; Grade VI, 92; Grade VII, 76; Grade VIII, 81. The coefficients of reliability for correct principle were as follows: .530, .630, .645, and .723; for correct answer they were .518, .528, .576, and .707. Using Brown's formula the coefficients of reliability for Gray's Silent Reading Tests were computed for thirty grade groups. These coefficients ranged from .55 to .85 with the median at .67. The number of pupils per group was less than 100 in only five cases. For several grade groups reliability coefficients were secured for Monroe's Standardized Silent Reading Tests which ranged from .222 to .907 with an average of .669.

Haggerty has computed the reliability for both Sigma 1 and Sigma 3 of his Reading Examination by having the same test repeated. In the case of Sigma 1 the interval between the two applications of the test was six weeks. For 200 children in Grades I to III inclusive the coefficient of reliability .84 was obtained. In the

case of Sigma 3 the interval between the two applications was only two days. For 126 pupils from Grades V to VIII, inclusive, the coefficient of reliability was found to be .885. For the sentence test alone the reliability coefficient was .769 and for the paragraph test, .806. For Thorndike's Scale Alpha for the Understanding of Sentences, McCall has reported a coefficient of reliability of .37. This was obtained by using a test similar to Alpha but not considered a duplicate form. Gates[5] reported reliability coefficients for Thorndike-McCall Reading Scale which ranged from .25 to .72. All of these were for pupils belonging to a single grade. For the Burgess Picture Supplement Scale the author has given coefficients of reliability ranging from .62 to .88 for grade groups from the second to sixth grades inclusive. In each case the number of pupils was relatively small. Gates gave coefficients of .62, .59 and .66 for three grade groups.

For the Otis Self-Administering Test of Mental Ability the author has reported an average reliability coefficient of .921 for the higher examination and of .948 for the intermediate examination. Presumably these coefficients are based on the scores secured from pupils for a sequence of several grades. For the separate tests of the Stanford Achievement Test the authors reported coefficients of reliability based upon separate grade groups which ranged from .75 to .96. When the composite score of all the tests was used the reliability coefficient was .98.

The relative reliability of written examinations and standardized educational tests. The data which have just been submitted indicate that the difference between the reliability of the two types of instruments is not as great as is commonly believed. The median of the reliability coefficient for written examinations given in Table IV is .65. There are many reliability coefficients for standardized tests in Table V which are less than this. Furthermore, the additional citations of coefficients of correlation in the above paragraphs indicate that for a number of standardized educational tests which have been very widely used the median of the reliability coefficients for grade groups is in the neighborhood of .65. Thus the conclusion seems justified that altho some of our more elaborate standardized tests, such as the Stanford Achieve-

---

[5]Gates, Arthur I. "An experimental statistical study of reading tests," Journal of Educational Psychology, 12:379, October, 1921.

ment Test, the Illinois General Intelligence Scale, and the Otis Self-Administering Test of Mental Ability, may be expected to yield measures whose reliability is greatly in excess of that of typical written examinations, many widely used standardized educational tests yield measures which possess about the same degree of reliability as the grades obtained from written examinations prepared by teachers and other school officials. It should be noted that reliability refers only to the variable errors of measurement. The constant errors as we shall show, (p. 40) are likely to be very much larger in examination grades than in the scores yielded by standardized educational tests. It should also be noted that the time required to give many of the standardized tests is much less than that devoted to a typical written examination.

The absolute reliability of examination grades. The statement that the reliability of a typical examination is equivalent to that of many standardized tests and only slightly less than that of a number of others still leaves a doubt with reference to the absolute reliability. For practical purposes the reliability coefficient of .65 needs to be interpreted in terms of the variable errors of measurement to be expected. The correlation tables for eight groups having a reliability coefficient of approximately .65 were taken and the scores translated into a five point system of school grades. It is assumed that these classes were typical and the highest scores were translated into a mark of "A," the lowest into a mark of "E." This was done in an arbitrary way but the results indicate roughly one meaning which may be attached to a reliability coefficient of .65. It was found that in 40 percent of the cases the students received the same grade in the two examinations. In an additional 42 percent the grade which they received on the first examination was only one point higher or lower than that received on the second. For example, if a student in this group made a "D" on one examination, he made an "E" or "C" on the other. The two grades received by the remaining 18 percent differed by two points or more.

Conditions tending to produce variable errors of measurement in examination grades. Several sets of examination papers were examined in order to ascertain the conditions which tended to produce the lowest coefficients of reliability and hence the largest variable errors of measurement. The most potent cause

appeared to be that the two teachers recognized widely different educational objectives in making out the two sets of examination questions. This seemed to be the case in Group 42, arithmetic, for which the coefficient of correlation was −.18. In Group 22, geometry, there was a difference in the general plan of the examinations; one teacher permitted the students to choose one of two questions in part of the examination while the other required that all questions be answered. This difference in the plan of the examination appeared to increase the variable errors of measurement. There was also a difference in the educational objectives recognized in that one teacher placed much more emphasis upon the practical application of geometry than the other.

Another cause which operated to lower the degree of correlation and hence to increase the magnitude of the variable error was the adherence to different standards of excellence by the teachers who graded the papers. For example, in Group 45, arithmetic, one teacher considered only the final answer to the exercise; if that was right the student received full credit—if wrong, no credit was given. The other teacher gave credit for correct principle. The coefficient of reliability for this group was .06.

It was noticed that in general pupils made higher grades on the tests set by their own teacher than on those set by another person. This appeared to be true even when distinct differences could not be identified either in the educational objectives or in the methods of grading of the two teachers. In Group 32 for which a reliability coefficient of .30 was obtained when the grades made on the first examination were correlated with those made on the second examination, a second coefficient was calculated by comparing the student's grade made on the examination prepared by his own teacher with that set by another teacher. This procedure gave a coefficient of correlation of .40. When the two classes were taken separately coefficients of .57 and .44 were obtained. These data tend to supplement the evidence already cited that differences in the content of the examination and in the plan of marking are potent factors in producing the variable errors of measurement.

The magnitude of constant errors in examination grades. It is probable that most of the teachers marking the examination papers did not recognize the distinction between "scores" and "grades"[6] and that the marks placed upon the papers were con-

[6]See page 11 for a statement of this distinction.

TABLE VI. DISTRIBUTION OF DIFFERENCES BETWEEN AVERAGES OF EXAMINATION GRADES

Difference	Frequency
50	1
32	1
29	1
27	1
22	1
21	1
20	2
18	3
16	1
15	1
14	2
13	2
12	1
10	2
9	1
8	4
7	3
6	6
5	4
4	6
3	5
2	6
1	8
0	3
Total.........	66
Median.......	6.2

sidered as "grades." In several instances the "grades" made on one examination were on the average much higher than those made on the other. If "scores" were used as "grades," any differences between the averages of the two sets of measures indicate the presence of constant errors. In order to secure an index of their magnitude the differences were calculated for the sixty-six groups to which two examinations were given. These are assembled in Table VI. For three of these groups the difference between the averages of the two sets of "grades" was zero; for eight other groups it was one. At the other extreme we find a difference of 50 in the case of one group. The median difference is 6.2.

It should be noted that the differences between the averages of two sets of examination grades are not constant errors. They are merely indicative of the presence of constant errors. If one examination was easy and the other hard the difference would be the sum of a positive error and a negative error. If both examinations were hard the difference would be smaller than the constant error in either average. The large differences shown in Table VI

are probably caused by the combination of an easy examination with a difficult one. This was very obviously true in the case of the one difference of 50. Furthermore, in interpreting Table VI it should be remembered that possibly some of the teachers recognized the distinction between "scores" and "grades," and the marks would have been appropriately adjusted before being used as "grades."

So far as it was possible to ascertain from an analysis of the examination papers the large differences are due to two causes—differences in the difficulty of the two sets of examination questions and in the severity of the grading. For example one of the examinations which produced a difference of 40 consisted of seven questions of which the pupils were permitted to answer any five. These questions were relatively easy. In the other examination, there were ten questions and the pupils were required to answer all of them. Very few were able to complete this second examination in the time allowed and the teacher appears to have counted the unfinished exercises as failures. Nine out of twenty-two children in the second group made zero on the examination. In this way a very large constant error was introduced but the coefficient of reliability for this group was .48.

Relative magnitude of constant errors in examination grades and in standardized test scores. In another place[7] the writer has discussed the magnitude of the constant errors in educational tests. In cases where there has been coaching for tests, intentional or not, or disregard for standard directions, large constant errors have been introduced. In one extreme instance a constant error of over three and a half years occurred in the mental age scores of a group of children. In general, however, because of the standard directions for administering the tests and scoring the papers, of the objectivity of the marking, and of the norms for interpreting test scores, the constant errors in standardized tests are very much smaller, and are likely always to be smaller than those found in examinations given by teachers. However, some reduction in the magnitude of the constant errors in examination scores will result when the use of either very easy or very difficult sets of questions is avoided and when a conservative plan of marking is followed.

[7]Monroe, Walter S. "The constant and variable errors of educational measurements." University of Illinois Bulletin, Vol. 21, No. 10, Bureau of Educational Research Bulletin No. 15. Urbana: University of Illinois, 1923, p.19-20.

Explanation of the apparent contradiction between the results of this investigation and previous studies of examination grades. The results of this investigation have caused the writers to revise their estimate of the accuracy of examination grades. The findings indicate that the errors are much less than they appeared to be from evidence resulting from investigations of the Starch-Elliott type. One naturally asks the question, "Why this apparent contradiction?" Starch and Elliott obtained similar results for three different examination papers and numerous other investigators have corroborated their findings. The mass of evidence accumulated is so extensive and uniform in character that one would naturally be inclined to accept the conclusions indicated in preference to the apparent contradictory results of the present investigation. However, a careful analysis of the procedures reveals that the results are not necessarily contradictory. The method followed by Starch and Elliott combines both constant errors and variable errors. The "grades" assigned to the examination paper in geometry were influenced both by the subjectivity of the marking and by the tendency of some teachers to grade high and of others to grade low. The present investigation has separated the variable errors from the constant. It has also shown that the examination scores have in some cases involved relatively large constant errors. The extreme differences between the grades assigned to the same paper reported by Starch and Elliott (see page 9) are easily explained when it is understood that they represent the combination of variable errors and constant errors. Especially is this true when we realize that the constant errors would likely be larger for teachers of different schools as in their investigation than for teachers in the same school as in the present investigation.

Conclusion with reference to relative accuracy of examination grades and scores yielded by standardized tests. As already indicated the writers believe that the data presented in this chapter show that examination grades are more accurate measures of achievement than many persons have considered them to be. Standardized tests yield scores involving errors, both constant and variable, but in the case of our best standardized tests these errors are distinctly less than the corresponding errors in examination grades. Furthermore, measurement by means of standardized tests usually requires much less time than is com-

monly devoted to written examinations. This conclusion refers to written examinations of the traditional type and administered under typical conditions. It is likely that written examinations and their administration may be improved so that the difference in the accuracy of examination grades and test scores will become much less than at present.[8]

[8]The conditions of standardized tests would have been more closely approximated if both sets of examination questions had been prepared by the same person and marked by different persons. If this had been done it is reasonable to expect that the coefficients of reliability would have been somewhat higher and the differences in the averages of the two sets of scores smaller.

CHAPTER V

THE CONTENT OF WRITTEN EXAMINATIONS

The data collected. In response to an invitation sent to superintendents and high-school principals in Illinois sets of examinations were received from fifty-six schools for the first semester and from fifty schools for the second semester of the school year 1921-22. A range of approximately sixty subjects was represented. It seemed desirable to restrict this analysis of sets of questions to the thirteen subjects listed in Table VII. The number of sets of questions and also the total number of questions are given in this table.

Classification of questions according to type. After considerable experimentation a list of fifty types of questions as given below was formulated.

Aims
Analysis
Cause (give)
Classification
Comparison
Completion
Conjugation
Construction (a figure, study or statement)
Construction (give the)
Contrast (general)
Contrast (specific basis)
Correction
Criticism
Decision (choice or preference)
Declension
Definition
Description (characterization)
Diagram (illustrate by)
Discussion
Effect (give the)
Evaluation
Example (illustrate by)
Expansion
Explanation (tell why or how)
Facts (definite number)
Facts (indefinite number)
Factoring
How many (tell)
Law (give the)
Mathematical operations of addition, subtraction, multiplication, and division
Method
Outline
Parsing
Proof
Punctuation (capitalize and correct sentences)
Recall
Reduction to lowest terms
Relationships (give the)
Rule
Scanning
Simplification
Source
Substitution (values for letters)
Summary
Solving for unknown quantity
Syllabus
Translation (foreign language to English)
Translation (English to foreign language)
Use (give the)
Where (tell)

TABLE VII. NUMBER OF QUESTIONS AND SETS OF QUESTIONS EXAMINED

Subject	Sets	Questions
English I	80	721
English II	83	694
English III	79	721
Algebra I	80	731
Plane Geometry	80	636
Latin I	81	683
Latin II	76	539
Physics	76	795
General Science	62	789
Civics	59	560
American History	62	550
Domestic Science	42	392
Domestic Art	41	368
Total	901	7621

All questions for the thirteen subjects mentioned in Table VII were classified under some one of these types. This classification was made by Mr. Souders with the assistance of a single clerk working under his immediate direction. Altho any classification of this kind is necessarily subjective, a relatively high degree of uniformity has, we believe, been secured.

Summary of classification. Twenty-six of the fifty types of questions were represented in six or more of the thirteen subjects. The relative frequency of each is given in Table VIII. This classification of examination questions shows a high frequency of certain types and very little or no use of a number of other types. If we omit Latin, Algebra, and Plane Geometry in which the nature of the subject-matter restricts the kind of question asked, we find that 32 percent of all the questions require "explanation." The next most frequent type used, 21 percent, calls for a "definite number of facts."

Frequently all questions are considered as belonging to one of two groups, "thought questions" or "memory questions." Such a definite classification is not, however, always possible. The character of the mental process involved in answering depends upon the person replying as well as upon the form of the question asked. Those questions calling for definite facts are almost certain to be based upon memory; on the other hand, those requiring classification, evaluation, contrast, etc. are likely to demand

TABLE VIII. PERCENT OF RELATIVE FREQUENCY OF TYPES OF QUESTIONS IN THIRTEEN SCHOOL SUBJECTS

Type	English I	English II	English III	American History	Civics	Latin I	Latin II	Algebra I	Plane Geometry	General Science	Physics	Domestic Science	Domestic Art
1. Analysis	1	—	—	—	—	—	—					—	
2. Cause or effect	1	—	—	4	—		1			2	1	1	1
3. Classification	1	1	1	—	1	—	1		—	1	—	2	1
4. Completion	2	2	1	—		—	—	—	1	—	—	—	2
5. Construction	14	13	13	1	1	2	5	1	12	2	2	3	10
6. Contrast (Gen.)	—	—	—		—	—					—	—	—
7. Contrast (Spec.)	2	4	5	3	4	2	1	—		3	4	3	3
8. Correction	3	3	3				—	—				—	
9. Decision	2	1	—	—	1				—	—		1	—
10. Definition	5	6	4	1	6	—	1	2	5	4	11	2	2
11. Description	6	3	6	2	6		2			6	5	5	6
12. Diagram	—	—	—				—	—	1	5	6	1	3
13. Discussion	1	3	7	16	7		1			4	5	7	3
14. Effect	—	—	1	3	1		1			2	—	2	1
15. Evaluation	—	—	—	—	1			—				—	—
16. Example	5	4	3	1	2	3	2	2	1	5	4	1	3
17. Explanation	21	22	15	36	38	8	13	1	6	42	37	34	28
18. Facts (Def. no.)	18	20	20	21	20	9	7	1	4	17	10	25	29
19. Facts (Indef. no.)	1	1	2	3	2	—	—	—	—	2	1	3	2
20. How many				—	—	—		—	—		—		
21. Outline	2	2	4	5	4	—	1		—	1	—	2	3
22. Proof	—	—	—					2	56	—	1		—
23. Recall	8	7	6	3	2	1	3			1	—	2	—
24. Rule	1	1	—		3	2	3	4	2	—	2	1	1
25. Use	1	2	1	—	—	3	1				1	2	2
26. Where	1	1	—	—	—	—			1		—	—	1
All other types	5	3	5	—	1	68	60	86	11	1	12	2	
Total Questions	922	906	802	601	761	760	578	646	699	667	887	500	580
Percent of Memory	21	19	16	4	11	73	65	69	74	9	24	9	6
Percent of Thought	79	81	84	96	89	27	35	31	26	91	76	91	94

The dash (—) means a frequency of less than ½ of 1 percent.

thought on the part of most students. If, however, such classifications or evaluations have been made in a previous class exercise some students may easily remember the answers and, in such a case, a thought question for one student becomes a memory question for another. For the purpose of this study Types 10, 20, 22, 25, and 26 have been designated as probable memory questions, the remaining types as probable thought questions. The percent of each group is given in the last two lines of Table VIII. These percents can be considered as only a rough indication of the relative frequency of these two very general divisions.

In her investigation of "the question as a measure of efficiency in instruction," Dr. Stevens[1] attempted to determine the relative number of thought questions and memory questions asked by teachers in a single class period. The percents of memory questions for history, English and science were 83, 55, and 67 respectively. This relative frequency is much larger than indicated in Table VIII. The difference may be due to the fact that in the present investigation only written examination questions were considered, but it is altogether likely that it is indicative of a real change in the type of questions which teachers commonly ask of their students.

Relation of questions to educational objectives. The questions which teachers ask during class periods constitute a concrete expression of the educational objectives which they are day by day setting for their students. The questions of the final examinations should, therefore, be representative of the types of educational objectives set in the different school subjects.[2] The emphasis upon memory and some of the simpler types suggests a need for a modification in emphasis in most of the school subjects.

Quality of examination questions. Altho the writers have no objective evidence to present in regard to the quality of examination questions, those submitted for this study were in general considered good. Catch questions or those stated so that they would not be understood easily by students were very rare. Many questions were stated so that the grading of the answers was objective and would indicate that their form had been influenced by

[1]Stevens, Romiett. "The question as a measure of efficiency in instruction." Teachers College Contributions to Education, No. 48. New York: Teachers College, Columbia University, 1912.

[2]See page 55 for a further discussion of objectives in school subjects.

the exercises of standardized tests. This was especially true of those examinations having questions of the true-false or recognition type. In the judgment of the writers the criticism that teachers are inclined to ask misleading or catch questions is not a valid one in the case of the examinations studied in this investigation.

CHAPTER VI

THE IMPROVEMENT OF WRITTEN EXAMINATIONS

Altho we now have a number of standardized tests which are superior to written examinations, and we have reason to believe that they will be used even more extensively than at present, there is need to give attention to the improvement of written examinations. It does not appear likely that standardized tests will ever replace written examinations. The latter type of measuring instrument will probably continue to be the most frequently used means of measuring the achievements of school children.

Written examinations may be improved by correcting the faults which have been noted in the preceding chapter. In this chapter we shall consider four important improvements: (1) Reduction of constant errors; (2) Reduction of variable errors; (3) Securing a greater agreement of the content of examinations with recognized educational objectives; (4) Simplification of the administration of written examinations.

There is some overlapping between these improvements. For example, the magnitude of errors in measurement, particularly variable errors of measurement, will be reduced by securing a greater agreement between the content of the examination and recognized educational objectives. The devices for simplifying the administration of examinations also tend to make the results more accurate.

Causes of constant errors in examination grades. The fundamental cause of constant errors in examination grades, i.e., "high grades" or "low grades," is the failure to recognize the distinction between "scores" and "grades." (See page 11.) A pupil's grade tells his standing with reference to a norm, i.e., the passing mark. When no distinction is made between "scores" and "grades" this norm is subjective. Altho the passing mark may be defined numerically as 70 percent or 85 percent it is fixed in the case of a particular examination by the difficulty of the questions and by the severity of the marking of the papers. Pupils will receive "high grades" when the examination is easy or the plan of

marking is generous. They will receive "low grades" when the examination is hard and a severe plan of marking is followed. If the teacher makes no distinction between "scores" and "grades" he sets the norm for a particular examination when he makes out the questions and decides upon the plan of marking. He implicitly expresses the opinion that the pupil whose achievements are barely "passing" will make a grade of 70, or the passing mark adopted. He also implies that the pupil whose achievements are exceptionally high will make a high grade, i.e., a grade of 95 or between 95 and 100. Such expressions are merely subjective.

Since the failure to recognize the distinction between "scores" and "grades" is the cause of constant errors the plan for improvement is obvious. The papers should be marked in terms of "scores." These may be on the scale of 100 but this is not essential. In fact it will probably assist a teacher in keeping the distinction in mind if the scores are not on the scale of 100. After the papers have been marked the "scores" should be translated into "grades" by comparison with a norm in which the subjective elements are reduced to a minimum.

A standard average grade used as a norm. The simplest objective[1] norm is a standard average grade. This may be set arbitrarily but a more rational procedure would be to take the average of the grades given in a school on a particular subject during a period of several years.

The standard average grade defines the grade into which the average score of a typical class should be translated. For example, if the standard average grade is 85 and the average score in a particular class is 57 the grade corresponding to this score would be 85. In case the class is made up of poor students the average grade of the class should be below the standard average grade. If the class is unusually bright their average grade should be higher than the standard average. The translation of the average score into the appropriate corresponding grade furnishes a basis for the translation of the other scores of the group.

The procedure just outlined is necessarily crude. It is partially subjective because the determination of the general status of

[1]The adjective "objective" is not intended to indicate perfect objectivity or even as high a degree of objectivity as we have in the case of many standardized tests. As used here it means that the norm is distinctly less subjective than the norm commonly implied in the usual examination.

the class is left to the teacher. However, the teacher may use previous school records or the measures obtained from a standardized test to assist him in arriving at a partially objective estimate of the general status of the class. The use of a standard distribution instead of merely a standard average grade represents a more systematic procedure.

1. Decreasing the magnitude of constant errors by means of a standard distribution of grades. For several years a number of educators have been urging that teachers make the distributions of their grades conform to a standard shape, i.e., that a specified percent of the members of a typical class be given a grade of A, another specified percent a grade of E, and so on for each of the marks adopted by the school.[2] A number of distributions have been recommended. For a five point system of grades several authors have recommended the following distribution, 7, 24, 38, 24, 7. Other distributions which have been advocated are 7, 18, 50, 18, 7 and 5, 15, 60, 15, 5.

The essential feature of the plan is a specification of the percent of the students of a given group who are to receive each mark rather than the particular form of distribution used. There is much evidence which indicates that the distribution of achievements of an unselected group of students approximates the normal probability curve.[3] If we assume that true measures of the achievements of an unselected group of 100 or more are distributed normally this adjustment does not fix the percent who are to receive each grade. The normal probability curve may be divided in many ways, for example, it is possible to divide the curve so that there would be 50 percent of A's, 20 percent of B's, 10 percent of C's, 10 percent of D's, and 10 percent of E's. In such a distribution a grade of A would be given to all students above the average of the class. An appropriate meaning could be stated also for each of the grades. A distribution which is symmetrical has certain advantages and one of those mentioned in the preceding paragraph is to be preferred. The particular standard distribution to be used is a matter of policy which each school should determine. Some argue that different standard distributions be adopted for

---

[2]If grades are expressed in percents the corresponding intervals such as 95 to 100, 90 to 94, etc. would be used instead of A, B, C, etc.

[3]The normal probability curve is bell shaped and is symmetrical with the average or median as a center.

the different years of the high school, some advocate different standard distributions for different school subjects. It should be noted, however, that there are certain advantages in uniformity. It would be desirable for all high schools, particularly those in a given state, to agree upon a common standard distribution and to use this for all subjects. Grades assigned in different schools can have a common meaning only when they conform to the same standard distribution.

The proposal that teachers make the distributions of their grades conform to a standard shape has met with much criticism. As in any controversy there have been extremists on both sides and many of those participating have given evidence that they failed to understand clearly the nature of the proposal of its essential features. Among the advocates of the use of a standard distribution are those who have insisted that the normal probability curve explicitly defines the students who must receive A's, who must receive B's, etc. Cases have been reported of instructors who frankly admitted that a certain student deserved to receive an A but that they had used up all the A's which the distribution allowed, and that, therefore, the student must be satisfied with a B. One hears also of instructors who announce at the beginning of a course that a certain number of the class must fail. It is rumored that in some of these instances the students enrolled have hired certain other students who were indifferent to their scholastic standing to enter the course in order to provide the requisite number of failures. The opponents of the use of a standard distribution have contended that there was no a priori reason why any student should fail and that always the quality of the student's work should determine his scholastic standing. Furthermore, they have pointed out that in any group of students brought together for instructional purposes it is extremely unlikely that the distribution of achievement would approximate at all closely their predetermined standard distribution. The mechanical and unintelligent application of a standard distribution by some instructors has given the opponents of the plan concrete examples of what they imagined to be the normal result of its use.

A standard distribution is merely a device which teachers may use in order to reduce to a minimum the constant errors in their grades, but to be helpful it must be used intelligently. It must be remembered that a standard distribution is a means and

not an end. Whenever common sense indicates that the distribution of grades for a particular class should depart from the normal distribution no instructor should hesitate to award the grades which he believes the students deserve. It is intended that the standard distribution will be closely approximated only for a large unselected group of students. A particular class very frequently is made up of a selected group of students. Furthermore, classes of the usual size, 20 to 35, are so small that frequently there will be significant departures from this standard distribution.

Translating scores into school marks by means of a standard distribution. A standard distribution is useful in translating examination "scores" into "grades." The examination papers should be marked in terms of a score. This score may or may not be on the scale of 100 points. In order to avoid confusion between "scores" and "grades" it is wise to use a scale of points shown so that the maximum score will not be 100. If the papers have been marked in this way the scores may be arranged in columns as indicated in the left hand margin. The first step in translating these scores into grades is to determine whether or not the class is typical. If an experienced teacher has had a class for several weeks he will usually be able to estimate its general status with a fair degree of accuracy. At the beginning of a school year or in the case of an inexperienced teacher some outside information is needed. The previous school record of the students may be studied but in many cases it will be more convenient to administer a general intelligence test. The average mental age and the distribution of the I.Q.'s of the class will be a very reliable index of the composition of the group. If the median I.Q. is distinctly below 100 the teacher may know that he has poor pupil material. If it is much above 100 he knows that the class consists of pupils better than the average. If there is an unusually high number of low I.Q.'s he may expect a relatively high number of low grades.

74 73 70	A
69 68 64 63 60	B
58 57 56 55 54 51 50	C
47 46 42 41 40	D
38 35 32	E

With the general status of the class in mind the scores may be grouped in conformity with the system of marks used. In the illustration in the left hand margin it has been assumed that the class is approximately typical. The percent of A's and also the percent of failures

are somewhat larger than the percent specified in most standard distributions. If the scores are arranged in the form shown below the general shape of the distribution will be more obvious. However, in the majority of cases it will be sufficient to use the arrangement given in the margin.

		58		
		57		
	47	56	69	
	46	55	68	
38	42	54	64	74
35	41	51	63	73
32	40	50	60	70
E	D	C	B	A

It is seldom that one will have exactly a symmetrical distribution of grades for a class of this size. Some departures from the standard distribution must be expected. In case the class is not typical one should expect marked departure from the standard distribution. For example, the distribution for a given class might be as shown below. In this there are no grades below passing but there are a number of poor students just above the passing mark. Also the percent of A's and B's is unusually large. Such a distribution is not normal but might well represent the distribution of grades for a particular class even when the normal distribution had been adopted as the standard. If the teacher is able to show that the general status of the class justifies such a departure he deserves commendation rather than criticism for his distribution.

			64	
	42		63	
	41	54	60	74
	40	51	58	73
	38	50	57	70
	35	47	56	69
	32	46	55	68
E	D	C	B	A

An accumulative distribution used as a check upon constant errors. A standard distribution is also useful as a check upon the grades given by a teacher over a period of several terms. When the grades for the entire period are assembled in such a distribution any general tendency on the part of the teacher to give too high or too low grades will be revealed. Each teacher should keep an

accumulative distribution of the grades in each subject he teaches. For example, a teacher in mathematics should keep an accumulative distribution of the grades given in classes of first-year algebra. When the total number of grades becomes large a comparison of this distribution with the standard distribution will reveal any tendency on the part of the teacher to grade too high or too low in this subject. A teacher should then take steps to correct any marked departures from the practise defined by the standard distribution. In large schools where there are several sections of the same subject it will be helpful to secure a distribution of grades each time they are issued. Any marked departures from the standard distribution will then be called to the attention of the teachers. However, one should avoid giving the impression that there must be uniformity with the standard distribution. Departures from this standard distribution are justified when the group of pupils can be shown to be selected. Thus a departure from the standard distribution is a cause for an investigation on the part of the teachers concerned. If evidence can be produced which justifies the departure no change in the system of grading should be used. On the other hand when investigation reveals no reasons why there should be departures from the standard distribution, the teachers should be urged to modify their system of grading so that a greater uniformity will be secured.

2. **Decreasing the variable errors in examination scores.** The reduction of the magnitude of variable errors of measurement in examination scores is to be secured mainly through the adoption of rules which will bring about greater uniformity in preparing and administering examinations. These rules should include specifications in regard to the effect of poor writing, poor spelling, and poor English upon a student's grade, and should be in agreement in regard to giving credit for correct principle and partial credit for exercises partly right or partly completed. The rules may properly include also specifications relating to the number and types of questions to be asked and the form in which they are to be presented to the students. For guidance in marking papers a teacher should write out, at least in abbreviated form, the correct answers to the questions. The accuracy of examination scores will be increased also by making the examinations more uniform with respect to content.[4]

[4]For recommended rules covering these and other points see Chapter VII.

It has been proposed that the use of types of questions which call for answers that may be objectively classified as either "right" or "wrong," would facilitate uniformity in marking the papers. This means of reducing the variable errors of measurement will be considered under the head of "simplifying the administration of written examinations."

3. Securing agreement of the content of examinations with recognized educational objectives. The intrinsic function of an examination is to measure certain achievements. In general the achievements for which we desire to secure measurements are those included in the recognized educational objectives. Hence, the questions should be in agreement with the objectives. Therefore, it is impossible to cover all details in a given subject-matter field. The questions should relate to the most significant facts, principles, etc. of the course. Catch questions and those calling for unimportant details have no place in an examination. For example, an examination in spelling should not include unusual or obsolete words, an examination in history should not call for obscure dates or other trivial facts.

In securing agreement the teacher should make use of such terms of minimum essentials as are available. For example, in spelling a teacher may very properly select the test words from Ayres' list of the one thousand most frequently used words or from some other carefully prepared minimum essential list. In geography a teacher will find the Hahn-Lackey Geography Scale a helpful source of questionings. In other subjects the teacher will not have access to terms of minimum essentials as complete as in these two subjects, but he should become familiar with curriculum studies and other investigations[5] relating to educational objectives.

---

[5]The following list is suggestive of studies relating to educational objectives:

Yearbooks of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company.

Part I of 14th—reading, writing, spelling, language and grammar, arithmetic, history, literature, geography.

Part I of 16th—reading, writing, spelling, arithmetic, history, physical education.

Part I of 17th—arithmetic, geography, reading, English, civics, history.

Part II of 17th—history, civics, economics, sociology, geography.

Part I of 19th—on new materials of instruction, reading, history, geography, mathematics, nature study, civics.

Part I of 20th—on materials of instruction—all subjects in elementary schools.

Part II of 22nd—the social studies in the elementary and secondary school.

"Arithmetic, course of study for the elementary schools, including the kindergarten and the first six grades." Course of Study Monographs, Elementary Schools, No. 1 Berkeley, California: Public Schools, 1921. 86p. (*Concluded on p. 56.*)

In some subjects there are valuable committee reports which give the consensus of opinion concerning the relative importance of the numerous distributions.

The teacher must assume most of the responsibility for securing the agreement between the content of the examination and educational objectives. In many of the high-school subjects he can obtain little assistance from such sources as just indicated. However, if this purpose is kept in mind and if he is really familiar with the subject which he is teaching, gross inconsistencies with recognized educational objectives will be avoided.

4. Simplifying the administration of written examinations. The administration of written examinations, particularly the marking of the papers, can be greatly simplified by the use of certain types of exercises. For example, in the true-false type of exercise the pupil merely indicates whether the statement is true or false. Instead of asking the question, "Why did the Puritans come to America in the seventeenth century?" we may ask whether the following statement is true or false. "The Puritans came to America in the seventeenth century seeking wealth." The pupil may give his answer to this exercise by writing a plus sign after the statement if he considers it true and a minus sign if he considers it false. In case the statement is dictated to him he may write after the number of the exercise the word "true" or "false" or the appropriate sign. The answering of such exercises requires very little of the pupil's time and the scoring is exceedingly simple. Questions which can be answered merely by "yes" or "no" also simplify the administration of examinations. Similar results can be secured with recognition exercises such as have been used in a number of standardized silent reading tests. The following is an exercise of this type.

Ayres, L. P. "A measuring scale for ability in spelling." N. Y.: Division of Education, Russell Sage Foundation, 1915. 58p.

Ayres, L. P. "Measuring scale for handwriting." N. Y.: Division of Education, Russell Sage Foundation, 1920. (Folder, chart.)

Bagley, W. C. and Rugg, H. O. "The content of American history as taught in the seventh and eighth grades." University of Illinois Bulletin, Vol. 13, No. 51. Urbana: University of Illinois.

Charters, W. W. Curriculum Construction. N. Y.: Macmillan Co., 1923. 352p.

Charters, W. W. and Miller, Edith. "A course in grammar." University of Missouri Bulletin, Vol. I, Education Series 9. Columbus: University of Missouri, 1915.

Hahn, H. H. Hahn-Lackey Geography Scale. Wayne, Nebraska: H. H. Hahn, State Normal School.

Hahn, H. H. Scale for Measuring Ability of Children in History. Wayne, Nebraska: H. H. Hahn, State Normal School.

Moore, E. C. Minimum Course of Study. N. Y.: Macmillan, 1923. 402p.

"The first president of the United States was: Christopher Columbus, Benjamin Franklin, George Washington, Thomas Jefferson."

In answering this exercise the pupil is asked to underline or mark in some other way the name required to make a true sentence. Completion exercises in which pupils are asked to supply words which have been omitted furnish still another means of simplification.

Directions for constructing a true-false examination.[6] 1. In constructing true-false exercises, a list of statements covering in some detail the portion of the subject on which the pupils are to be examined should be prepared. Some of the statements can then easily be changed so that they are false. The untruth of a statement should not be too obvious or it will be worthless for testing. Also statements should be selected which require an acquaintance with the subject in order to determine their truth or falsity.

2. In a true-false examination the number of true statements should approximate the number of false statements, and the arrangement should be such that there is no regular sequence between true statements and false statements.

3. Since the pupil can give his responses very quickly, the examination should consist of not less than fifty statements. A true-false examination of one hundred statements can be given in the time usually devoted to an ordinary examination.

4. The examination should be mimeographed or printed so that each pupil will have a copy. He may give his answers in the margins of the sheets, or, if it is desired to use the same set of papers with another group of pupils, he may be given a sheet of paper on which there are numbered blanks. The pupils will then be asked to record in the blanks their answers to the corresponding exercises. A less desirable plan, which may be followed when it is not possible to secure mimeographed copies of the examination, is to read the statements to the pupils and have them record their answers in numbered blanks. The disadvantage of this plan is that the pupils do not have a satisfactory opportunity to study the statements. Also the class may give some indication of the answer if a statement appeals to them as being ridiculous.

5. The pupils should be given specific directions in regard to answering exercises about which they are uncertain. One writer

[6]For an example of a true-false examination, see Appendix p. 69.

has suggested that the pupils be instructed to guess concerning the truth or falsity of the statement. Another writer who has used this type of examination instructed the pupils as follows: "First, go through the list quickly and mark all that you know for certain, then go back and study out the harder ones. Do not guess; the chances are against you on guessing. Don't endanger your score by gambling on those questions about which you know nothing." This second procedure is probably the better.

The scoring of a true-false examination. Since only two responses are possible, it is obvious that a pupil may give a correct response as the result of chance. In order to take this possibility into account, a pupil's score on an examination of this type is the number of exercises answered correctly minus the number answered incorrectly. Exercises not attempted are not counted.

Directions for constructing a recognition examination.[7] In constructing this type of examination none of the proposed answers should be too obviously incorrect. An exercise can yield an indication of a pupil's achievement only when he is forced to use judgment in determining which of the proposed answers is suitable. For example, the illustrative exercise given would be practically worthless for testing purposes if all the names, except that of George Washington, were of persons living today or of persons having no connection with our national life. In applying this type of exercise to the field of arithmetic the proposed answers should include erroneous answers which pupils are inclined to give: if the exercise called for the quotient of two fractions, one of the proposed answers should be the product of the fractions and another their sum, and perhaps another should be the fraction obtained by taking the sum of the numerators as a new numerator and the sum of the denominators for a new denominator. When the correct answer is included in a group of such answers as these, the pupil who does not know how to find the quotient of such fractions will be unable to determine the correct answer except as a matter of chance. On the other hand, if all of the answers except the correct one were integers or were so large that they were obviously incorrect, a bright pupil who knew nothing about division of fractions would be able to select the correct answer. The correct answer should not always be found in the same position; sometimes it should be

[7]For an example of a recognition examination see Appendix p. 75.

first, sometimes last, and sometimes in an intermediate position. As in the case of the true-false examination, a recognition examination should consist of a large number of exercises.

Examinations of this type should be mimeographed or printed and each pupil should have a copy. Definite instructions concerning methods of work should be given. It is probably best to instruct the pupil to work through the test rapidly, answering those exercises about which he is certain. He should then go back over the list and try the more difficult ones. Not fewer than four proposed answers should be included in each statement and the pupils may be instructed to guess if they do not know, since the chance of success by guessing is slight. The pupil's score on an examination of this type may be taken as the number of exercises done correctly.

A somewhat unusual but interesting type of recognition exercise is that described as a "matching contest." In this a pupil is given two lists of statements, the first numbered 1, 2, 3, 4, 5, etc., the second marked A, B, C, D, E, etc. In the second list, there is a statement which corresponds in meaning to a statement in the first list and the pupil is to pair these statements, marking by the number of the first list the letter of the corresponding statement of the second. For example, in the exercises given below: by the date marked (5), 1898, we place the letter B to indicate the event for which that date is significant. It is difficult to construct such examinations so that they will require reasoning on the part of the student. Their most important use is in the elementary school for rapid drill in certain phases of some subjects, such as definitions in geography and grammar, etc. The following exercises, selected from the Spokane United States History Test, illustrate the use of such an examination in linking a certain date or person with the corresponding event.

1. 1846	A. Lincoln's Emancipation Proclamation
2. 1865	B. Spanish-American War
3. 1863	C. Beginning of World War
4. 1917	D. Declaration of Independence
5. 1898	E. United States entered World War
6. 1789	F. Election of Washington as President
7. 1792	G. War with Mexico began
8. 1776	H. Invention of the cotton gin
9. 1861	I. Lee's surrender at Appomattox
10. 1914	J. Beginning of Civil War

1. Foch	A. Destroyed Spanish fleet in Manila Bay
2. Lincoln	B. Invented the telephone
3. Fulton	C. Leading Confederate General
4. Dewey	D. Wrote the Declaration of Independence
5. Pershing	E. Invented the steamboat
6. Bell	F. Commanded allied armies in the World War
7. Edison	G. Was President during the Civil War
8. Jefferson	H. Commanded American Forces in the World War
9. Lee	I. Was Revolutionary patriot, author, and inventor
10. Franklin	J. America's most famous inventor

Directions for constructing completion exercises.[8] A completion exercise should be constructed so that no suggestion will be given of the correct words to be written in the blanks. Furthermore, the facts to be supplied should be important. The best plan is to prepare a list of important statements and principles covering the portion of the subject over which the pupils are to be examined and then from these statements to strike out a certain significant word or phrase. In every case, if it is possible, the words omitted should be such that only one answer will be correct. Since little writing is required of the pupils they may be asked to fill in as many as one hundred blanks.

The scoring of completion exercises is not as highly objective as in the two types mentioned above. Pupils will tend to write a variety of words in the blanks. Different words may have almost the same meaning, and frequently the scorer will be compelled to determine whether the meaning of some word is sufficiently near that of the correct answer to justify giving the pupil credit for having answered the exercises correctly. However, by a careful selection of statements and of the omitted words, this subjectivity may be greatly minimized. For example, in the sentence, "The first Continental Congress was held in . . . .," only one possible word can be correct. In using completion exercises it is necessary to provide each pupil with a mimeographed or printed copy of the examination. The pupil's score is the number of blanks filled in correctly.

Advantages of the "new examination." Examinations consisting of exercises of the types described above have certain obvious advantages. There will be a large saving of time for both teacher and pupil. The pupil is called upon to do little or no

[8]For an example of a completion examination see Appendix p. 73.

writing in giving his answers and therefore is able to respond to a large number of exercises. The teacher in scoring will have little or no occasion to use judgment as he will need only to note the brief responses given by the pupils. Thus the labor of scoring will be greatly reduced and, more important, the scoring will be much more highly objective than that in the marking of examination papers of the usual type. The saving of time in the giving and scoring of the "new examination" will more than offset any additional time that may be expended in its construction. Another advantage is that the new examination can be made more comprehensive. Examinations as a rule consist of ten questions. Some are limited to a smaller number. Consequently the scope of examinations of the traditional type is necessarily narrow. "New examinations" of the true-false type should consist of not less than fifty exercises and may have as many as one hundred. Other types of the "new examination" should be of a corresponding length. Hence a "new examination" will usually be more comprehensive than a traditional examination.

Limitations of the "new examination." There are certain limitations of the new examination which should be noted along with its advantages. It can not be used in mathematics except to a limited extent. It can not be used at all in English Composition. In other subjects there are many phases of achievement which are not measured directly by examinations made up of exercises of the types described above. Hence, altho the "new examination" is more comprehensive with reference to information, and does measure certain types of achievements, it is likely that pupils would miss much valuable experience and training if they were not at times asked to "compare," "explain," "discuss," "define," or "tell why." They should also be asked to summarize material presented on a topic or to apply certain principles. The following questions taken from Hahn's Scale for Measuring Ability of Children in History appear to require mental processes distinctly different from those called for by the "new examination."

1. "State points of similarity between the position of the United States in 1812 and their position in 1912."

2. "Arrange the following events in order of cause and effect: Force Bill, Carpet Baggers, Fifteenth Amendment, Negro Rule in Some of the Southern States, Ku Klux Klan."

3. "Name the presidents of the United States since 1892."

An intelligent attitude toward the "new examination." The simple administration of the new examination and other attractive features should not blind one to the limitations just mentioned. As indicated in Chapter II written examinations do more than merely secure measures of achievement. If they consist of the right kind of exercises they afford significant educational opportunities. The educational opportunities of the "new examination" are necessarily restricted, and it would be unfortunate if it entirely replaced examinations of the traditional type. The new examination, however, has a place. It may be used occasionally in most school subjects. It is useful when a teacher wishes to test the acquaintance of a class with a wide range of facts. It has little diagnostic value and examinations of the traditional type should be used when information is desired concerning the weaknesses of different members of a class. For this reason the "new examination" is more appropriate for use at the end of a term than for tests during the term which have as their purpose both measurement and diagnosis.

CHAPTER VII

RULES FOR THE PREPARATION AND ADMINISTRATION OF WRITTEN EXAMINATIONS.

Below, a group of suggested rules governing the preparation and administration of written examinations are given. These represent the opinion of the writers which is based upon a careful study of the problems involved, as well as upon several years of experience in the measurement of school achievement.

1. Final examinations should be required. In school subjects such as shop work, in which the performances secured from pupils are highly objective, the waiving of this requirement may be justified. When final examinations are given no student should be excused from them because of high daily grades, deportment or attendance. (See p. 16)

2. The content of final examinations should agree as closely as possible with recognized educational objectives. In fields where minimum essentials have been determined they should be used as a basis in formulating questions. (See p. 55)

3. The questions should be definite and stated so that all pupils will interpret them alike. Questions relating to items of minor importance should occupy a minor place in examinations. Questions relating to points which have not received attention in the course should be omitted.[1] (See also rule 11.)

4. When the necessary equipment is available the questions should be mimeographed or typewritten so that each student will have a copy on his desk. In case they are written on the board the teacher should make certain that all pupils are able to read them correctly. It is well in either case to read the questions aloud to the class.

5. The examination should be sufficiently difficult so that few pupils will make perfect scores. (This rule should not apply

[1]Frequently in an examination the difficulty of the question is due to the lack of emphasis placed upon it throughout the term because it deals with a relatively unimportant topic. Other topics, difficult in themselves, but emphasized because of their importance, furnish the easier questions.

when the purpose of the examination is to determine which students have attained a given standing which includes perfection of performance.) (See p. 12)

6. Usually the examination should be long enough so that every member of the class will be kept busy for the entire period. It is better to make the examination too long than to have it too short. In this way it becomes possible also to take into account the student's rate of work in determining his grade. Appropriate adjustments can be made in interpreting "scores" into school "marks." (See p. 22 and 13)

7. In questions asking for a discussion or explanation indicate the completeness of the discussion or the degree of elaborateness expected in the answer. (See p. 24)

8. Time may be economized for both students and teacher by using some form of the "new examination." This type of measuring instrument, however, possesses certain limitations which should be kept in mind. The exclusive use of it would be unwise. (See p. 61)

9. Unless the students have a definite understanding of the methods of work which are to be followed, the teacher should give them explicit directions concerning such matters as the order in which the questions are to be answered, the desired arrangement of the work, and any other items in which there is an opportunity for pupils to adopt different procedures. (In the case of most standardized educational tests, the directions to students are very detailed and explicit.) (See p. 24)

10. Approximately ninety minutes should be allowed for a final examination in most high-school subjects. The time which teachers should devote to the preparation of the questions will depend upon their experience and upon their practise during the semester. It is recommended that a teacher make a record throughout the term of questions which in his judgment are suitable for a final examination. From two to three hours should be sufficient for marking a set of twenty-five examination papers. If a teacher finds that a longer time is required, he should endeavor to modify his procedure so that this work can be done more quickly. (See p. 19)

11. Altho any weighting of questions by a teacher will be subjective, it is probably desirable to weight the questions, particularly in cases of extreme differences in value. The weighting,

however, should be upon the basis of social importance rather than upon mere difficulty. (See p. 22)

12. It is advisable for the teacher to write out at least in an abbreviated form the answers to examination questions before he begins marking the papers. In mathematics and other subjects where a definite answer is required and only one can be accepted as correct, the need for this rule is not as great as in such subjects as geography, history, literature and certain phases of science. However, a list of correct answers will usually mean a saving of time. (See p. 23)

13. Except in courses in English a pupil's grade should not be intentionally lowered for errors in spelling or for poor handwriting. As a rule the grade should not be lowered because of poor English, unless the quality of the English is evidence of unsatisfactory reasoning. Rules covering these points, as well as others concerning which teachers might differ, should be formulated by the principal in conference with his teachers or at least a committee of them. The rules thus formulated should be carefully followed by all of the teachers. (See p. 20)

14. In marking the papers more accurate results will in general be secured if the answers to one question are marked on all the papers before those for another question are taken up. (See p. 23) When it is desired to mark all of the questions on one paper before taking up another, the "sorting method" should be used. According to this procedure the papers as they are read are sorted into piles, the best ones being placed in the first pile, the next best in second pile, etc. Five distributions will, in most cases, prove sufficient. After all the papers have been distributed they should be reread, one pile at a time, and compared with each other. If these papers do not possess approximately the same value, changes in the sorting may be made. Grades may then be assigned to the papers in the different piles.

15. The distinction between "scores" and "grades" should be kept in mind. (See p. 11) The papers should be marked first in terms of scores. In doing this an appropriate number of points should be determined for each question. It is not necessary that the total of these points be 100. (See p. 52)

16. The point scores assigned to the examination papers should be translated into school marks. In doing this the use of a standard distribution will be found helpful, and will operate also to decrease the magnitude of the constant errors. (See p. 50)

APPENDIX

(Questionnaire Sent to High-School Teachers)

QUESTIONNAIRE RELATING TO THE USE OF WRITTEN EXAMINATIONS IN HIGH SCHOOLS

Teacher...High School...City..........................

The major subject which I am teaching is..

The following questions are to be answered only with reference to the major subject you are teaching.

1. Approximately how much time do you use in preparing questions for a final examination which the students are allowed a total of 90 minutes to answer?.. ....Min.

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester.)

2. In preparing a set of examination questions do you usually attempt to arrange these questions in order of ascending difficulty?............ Yes No*

3. Do you prepare in written form carefully worded directions to the students regarding the procedure they are to follow in answering the questions? (These directions might include such points as, order in which questions are to be answered, length of answers, arrangement of work, etc.) Yes No

4. Which of the following methods do you use in presenting questions to the student?..

(a) Writing the questions on the board.. Yes No

(b) Furnishing the pupils with a mimeographed, carbon or printed copy of the questions... Yes No

(c) Dictating the questions to the pupils..................................... Yes No

5. Is it your custom to make the examination long enough so that practically none of the students will answer all of the questions in the time allowed?... Yes No

---

*Underline the answer (Yes-No) which you desire to make.

6. Is it your custom to make the examination short enough so that practically all of the students will answer all of the questions in the time allowed?.. Yes No

7. If you have given an affirmative answer to Question 6, do you note the time each student spends in writing his answers?.................................. Yes No

8. What proportion of the final mark for the semester is based upon the final written examination grade?.. ....%

9. In assigning grades to examination papers do you attempt to have their distribution conform to any standard form such as the normal distribution?.. Yes No

10. Do you usually grade all the answers on one paper before taking up those of another paper?.. Yes No

11. Do you usually grade the answers to one question on all of the papers before taking up the answers to a second question?........................ Yes No

12. Instead of marking the answers to each question separately do you attempt to estimate the value of the paper as a whole?........................ Yes No

13. Before starting to grade a set of examination papers do you write out the answers which you consider correct?.. Yes No

14. When you consider the questions of an examination to be unequal in difficulty is it your practise to give more credit for a correct answer to a difficult question than for a correct answer to an easy question? Yes No

15. Approximately how much time do you use in marking the papers of a final examination which the students are allowed a total of 90 minutes for answering? Estimate as accurately as possible. Base this answer on a class of 25 students... ....Min.

16. In marking examination papers do you intentionally lower a student's mark in the case of

(a) poor writing... Yes No
(b) poor spelling... Yes No
(c) poor English... Yes No

17. In the case of questions which are essentially mathematical in character do you give credit to the student for using the correct principle even though the final answer be wrong?.. Yes No

NOTE: It is desired that only teachers of mathematics, physics, and chemistry answer No. 17.

NOTE TO TEACHER: When you have answered the above questions please return this questionnaire to your principal.

(Questionnaire Sent to High-School Principals)

Principal.......................................High...City............................

1. Do you require your teachers to give final examinations at the end of each semester?.. Yes No*

("Final examination" as used in this questionnaire means an examination which is given at the end of a semester and which is based on the work of the entire semester)

If a *negative* answer is given to the first question, no answer is expected for the remaining questions.

2. a. Is it the practise in your school to exempt certain students from final examinations?.. Yes No

b. If so, what are the conditions (requirements) upon which you base exemption?

1. Deportment.. Yes No

2. Scholarship.. Yes No

Other requirements.

3. How many minutes do you allow for final written examinations?.... ....Min.

4. Because final examinations have proven unreliable some educators urge that students be given more than one comprehensive examination in each subject, and that these examinations be given on different days. Do you require more than one such final written examination in each subject?.. Yes No

5. a. In marking examination papers is it the practise in your school for the teachers to subtract from a pupil's grade for

1. poor writing.. Yes No

2. poor spelling.. Yes No

3. poor English.. Yes No

b. Are your teachers accustomed to giving more credit for correct answers to difficult questions than for correct answers to easy questions?.. Yes No

c. Is it the practise of your teachers when computing a semester mark to add or deduct credit in proportion to the time used by a pupil in answering the final written examination questions? Yes No

6. a. Have you advised your teachers as to what proportion of the final mark for the semester should be based on the final written examination?.. Yes No

b. Have you made a definite requirement in this respect?........... Yes No

c. If so, what proportion of the final mark for the semester do you require to be based upon the final written examination mark?.. ....%

7. Additional information pertaining to this topic is called for in a a second questionnaire which is to be filled out by high school teachers. Would you be willing to distribute this questionnaire among your teachers? If so, will you kindly indicate the number of teachers in your high school?.. ....

---

*Underline the answer (Yes-No) which you desire to make.

EXAMPLES OF "NEW EXAMINATIONS"

(The following "new examinations" are given for purposes of illustration. They may include several exercises which will prove unsuitable when given to pupils.)

TRUE-FALSE EXAMINATION IN PHYSIOLOGY

Prepared by

Bureau of Educational Research

University of Illinois

Name...Boy or Girl....................................

Age last birthday........................Next birthday will be..19.......

Grade.......................Date...............................City..............................State..........................

School...Teacher...

Below you will find a number of statements. Some of these statements are true, others are not true. Read each statement carefully, then if it is true mark a plus (+) in the column to the right of the sentence. If the statement is not true mark a minus (−) in the column to the right.

EXAMPLES

Answers to be written here

Read the statement below very carefully.

1. **Fats will form a lasting mixture with water.**

This is not a true statement so you will place a minus (−) sign in the column. Now read the second sentence.

2. **The layer of fat just beneath the skin is more than one-tenth of an inch thick.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the third sentence.

3. **The union of oxygen with any substance produces heat.**

This is a true statement so you will mark a plus (+) sign in the column. Now read the fourth sentence.

4. **Nitrogen constitutes only one-fifth of the volume of the air.**

This is a false statement so you will mark a minus (−) sign in the column...

PHYSIOLOGY

Answers to be written here

1. The kidneys vary 12 inches to 16 inches in length.

2. The external poisoning of the skin by poison ivy or sumac never results seriously.

3. A person having a good mind must necessarily have a large brain.

4. Color blindness is more prominent in men than women.

5. Plenty of fluids should be drunk at the time of eating solid food.

6. Bones are composed of animal and mineral matter.

7. The nails are hardened outer skin or epidermis.

8. The use of alcohol increases the tendency to commit crime.

9. A full grown person contains about six quarts of blood.

10. The brain is almost perfectly spherical in shape.

11. The kidneys are almost perfectly round.

12. All animals are made up of cells.

13. Substances, like glass, which permit rays of light to pass through them readily are said to be opaque.

14. The sense organs of smell are located in the lining of the cavity of the nose.

15. The skin is composed of two layers of tissue.

16. The end organs for taste occur in the mucous membrane of the tongue.

17. To extinguish the burning clothing of a person, it is necessary to wrap him in something to exclude the air.

18. All of the interior of the spinal cord is filled with gray matter containing nerve cells.

19. The great difference in the complexion of persons is due largely to the pigment lying in the epidermis.

20. Cancer is caused by germs growing in the tissue.

21. Diphtheria can be controlled by the use of Diphtheria antitoxin.

Answers to be written here

22. The brain is separated into two parts or hemispheres by a great longitudinal fissure.

23. When oxygen is separated from other substances the process is called oxidation.

24. Infectious diseases are due to changed methods of work and growth on the part of cells in certain regions of the body.

25. The use of alcoholic beverages builds up the body and makes the muscles stronger.

26. The great majority of grown people have been infected with tuberculosis germs.

27. The sense organs are the terminations of the sensory nerves serving to carry impressions to the spinal cord or brain.

28. Farsightedness is often caused by a blow on the eye.

29. An antiseptic is a substance which merely restrains the germs from growing.

30. The brain is in communication with the rest of the body by means of nerves.

31. The cerebrum is the path of communication between the nerves supplying the arms, trunk, legs, and brain.

32. The chief function of muscles is to hold up the body.

33. All milk contains bacteria.

34. The alcohol used in drinks is produced by the growth of yeast in a liquid containing sugar.

35. Our blood contains white corpuscles which destroy disease germs.

36. More people die daily from diphtheria than from tuberculosis.

37. The use of tobacco increases the strength of the muscles.

38. The use of tobacco makes the nerve cells function more keenly.

39. The chewing of dry bread aids the digestion as much as the use of gum.

	Answers to be written here
40. Air is composed chiefly of two gases, oxygen and nitrogen.	
41. The first step in treating a person who has been poisoned is to give an emetic.	
42. Light is produced by waves of a substance called ether.	
43. Non-infectious diseases are caused by small plants or animals called parasites feeding upon the human body.	
44. Alcoholic beverages have great value in curing disease.	
45. A drink of alcoholic beverage in the winter time causes a man's body to become warm.	
46. Each portion of the brain has its own definite work to perform.	
47. Fainting is caused by an over-sufficient supply of blood being sent to the brain.	
48. The spinal cord may act independently of the brain and produce many of the muscular movements necessary in routine work.	
49. The germs of typhoid fever usually gain access to the body by being breathed in with air.	
50. Narcotics are substances which cause any organs of the body to act more vigorously than is their custom.	

Directions to teachers: After the four examples have been studied by the pupils, read the following directions to them: "On the next page you will find a number of statements similar to the ones you have just read. You are to place a plus sign or a minus sign in the column to the right of each statement just as has been done on the first page. Mark all of the statements that you are sure you can answer correctly. If you find a statement that you are not sure you can answer correctly, study it carefully and then mark the answer you think will be correct. If you find a statement you know nothing about, make no attempt to mark it, as guessing counts heavily against you. You will have 25 minutes for the test. I shall expect you to stop promptly and turn your folders face down on the desk when I tell you to do so. Ready-Go."

In computing the score of each pupil on a test subtract the total number of wrong answers from the total number of right answers. Such scores are called "point-scores." In interpreting them it is advisable to form a distribution which will show how many pupils received each score. From the distribution it is possible to work out a basis for translating the point scores into the usual kind of school marks.

NOTE: The "Directions to teachers" given above would not appear on the usual printed examinations. They are placed here for the convenience of teachers.

COMPLETION EXAMINATION IN AMERICAN GOVERNMENT

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or Girl...

Age last birthday.........................Next birthday will be..19........

Grade.........................Date.................................City...........................State...............................

School...Teacher...

Below you will find a number of statements. In each statement one or more important words have been omitted. Each blank in the sentence shows where a word has been left out. Read each statement carefully, then write in the blank the word which completes the meaning of the statement. You will be allowed 15 minutes for the test.

1. The primary purpose for which government exists is the.........................of our lives and property.
2. Citizenship may be acquired by.........................in this country or by a process of ....for natives of other lands.
3. Our national government derives its authority from the.........................of the United States through our national.................................
4. The legislative power granted to the national government is vested in a Congress of ....houses, the smaller of which is called the.................................and the larger the............................
5. The execution of the laws made by.........................is intrusted to the............................ of the United States.
6. All judges connected with the national courts are appointed for life with the consent of..
7. Most of the candidates for office which are filled by popular vote are nominated directly in.......................................
8. The Fifteenth Amendment of the United States Constitution prevents the states from denying to citizens the right to vote on account of.., ..., or previous condition of................................

9. Practically all of our law-making bodies are made up of................................chosen for short terms from........................into which the states, counties and cities are divided.

10. The first permanent English settlements in America were made in what is now the state of...

11. In a county, the records of the county board and other official papers are preserved by the county..

12. All cities are public corporations created under............................municipal laws.

13. Every incorporated city obtains from the........................government a...................... under which it may elect its officials and conduct its business.

14. Civil service employees may be removed from service only for................................

15. The power of impeaching a state officer is given to the...

16. The................................is by far the most prominent and powerful executive official in the state. Very........................state officers are appointed by him or are responsible to him.

17. All important officials connected with the executive or judicial service of the United States may be removed by................................through the lower house of Congress and by................................in the senate.

18. Far more property is destroyed by............................than by all other agencies.

19. There is no task of state and local government which outranks in importance that of providing an..education at public expense.

20. All rivers and canals within a single state are controlled by the............................... in which they are located.

21. Most of the revenue for state and local governments is secured by a........................ on............................

22. A state........................is the fundamental law which the people of the state have arranged for their government and protection.

23. A state constitution can be changed by means of an..

24. The three-fifths compromise provided that five........................should be counted as equal to three........................when reckoning the................................for either direct taxation or representation.

RECOGNITION EXAMINATION IN ALGEBRA

Prepared by
Bureau of Educational Research
University of Illinois

Name..Boy or Girl...

Age last birthday........................Next birthday will be...19........

Grade........................Date........................City...........................State..

School..Teacher..

Below you will find a number of statements. In each statement a word or number has been omitted. At the close of the statement several words or numbers have been given. One of these is the correct answer. Select the word or number which you think is correct and draw a line under it. Most, if not all, of the examples can be solved by mental calculation. If any figuring is necessary, work on the margin of the page. You will be allowed 17 minutes for the test.

1. Numbers that are represented by letters are called............................numbers.
   substituted—literal

2. When two or more letters are multiplied together each is called a...............................
   of the product. factor—coefficient

3. If a man rides a certain distance in 10 hours, in h hours he rides...............................
   $10h$; $\frac{h}{10}$; $\frac{10}{h}$

4. The statement $2x + 5 = 29$ is called an........................ identity—equation.

5. If 16 is subtracted from three times a certain number the result is 110. The number is........................ $36\frac{2}{3}$; $31\frac{1}{3}$; 42

6. A number which is a factor of two or more numbers is called a........................factor.
   common—equal

7. If there are two equal factors of a number, either is called the........................of the number. square root—common factor

8. To multiply algebraic fractions take the................................of the numerators for a new numerator and the product of the denominators for a new denominator.
   sum—product

9. A fraction whose numerator or denominator (or both) contains fractions is called a ....fraction. multiple—complex

10. A......................is a statement of a fact which is to be proved. theorem—axiom

11. The name given the + sign is........................... negative—positive

12. To find the sum of two numbers whose signs are opposite, take their....................... regarding each as positive, and prefix the sign of the larger number to the answer. sum—difference—product

13. Whenever a number occurs without a sign, the...........................sign is to be understood. $\times$; $+$; $-$

14. The number denoting the power of a term is called the.. prefix—exponent

15. If $a = 2$, $b = -3$, and $c = -5$ then $\frac{2abc}{a} =$ .... -6; -2; 30

16. In adding like terms add the coefficients for the new coefficient and....................... it by the common factor. multiply—divide

17. An expression which contains more than one term is called a.................................. monomial—polynomial

18. If the length of a rectangle = 4 feet more than twice the width, the perimeter = 56 feet. The length = ...feet. 8—12—16

19. Any term may be transposed from one side of an equation to the other, provided its ....is changed. sign—value

20. Any equation which contains no higher power of the unknown letter than the first is called a................................equation. radical—simple

21. The exponent of the product of two powers of the same number is equal to the ....of the exponents of the factors. product—sum

22. To raise the product of two numbers to any power, raise the numbers separately to that power and take their........................ product—sum

23. The square of any two numbers is equal to the square of the first number............... twice the product of the two plus the square of the second number. plus—minus

24. A 20 foot ladder rests against a building, the bottom of the ladder being 12 feet from the cellar wall. The top is................feet from the ground. 8-16

25. In division, the sign of the quotient is........................whenever the dividend and divisor have like signs. $-$; $+$

26. In finding the quotient of two powers of the same number the exponent of the quotient is equal to the exponent of the dividend....................by that of the divisor. increased—diminished

27. $(3x^2-2x-1)\div(x-1)=$........................ $3x+1; 3x-1$

28. A factor which has no factor except itself and unity is called a......................factor. prime—multiple

29. The product of all the common prime factors of two or more numbers or expressions is called their............................common factor. highest—lowest

30. If one number is exactly divisible by another, the first is called a....................of the second. divisor—multiple

31. In algebraic fractions the dividend is called the....................................
denominator—numerator